

nature

An abstract illustration of a DNA double helix structure, rendered in white and grey, winding across the cover. It is surrounded by various molecular models, including a green textured sphere, a blue and white complex structure, and a red beaded chain. Numerous glowing yellow and orange particles are scattered in the background, creating a sense of dynamic activity.

ISLAMIC SCIENCE

Beyond the
blame game

UNIVERSITY

Safety
in numbers

CONVENTIONAL

MEDICINE

Can modern
science crack it?

ENVIRONMENTAL

Science
has

SUPERCOILED

Antitumour 'topo' inhibitor slows
DNA supercoil removal

A geographical error

The most suitably qualified person should secure an important post, in almost all instances. But competing interests among nations mean that this is not always the case in Europe.

The poster most likely to be found hanging in the offices of large European organizations is not a picturesque Alpine view, but rather a version of a much-loved cartoon known as 'European hell', in which the British are the cooks, the French are the mechanics, the Germans are the lovers, and everything is organized by the Italians. This gentle mockery of national stereotypes conceals a much more true-to-life hell of the European bodies themselves, as they try to function efficiently while being politically beholden to national interests. Too often, key appointments in European organizations, including those related to science, are made on the basis of nationality rather than merit.

This sometimes means that the most competent person to head a department will be passed over if a compatriot happens to head an unrelated department in the same organization. It is usual for countries to put forward their own candidates and although few of these will be outright incompetent, more than a few will owe their candidature more to political circumstance than to expertise. So the right person for a particular job might be the wrong nationality. And the candidate with the right nationality may well be the wrong person.

Take, for example, the European Space Agency, which now has 17 member states. It offers nine directorships, plus the position of director-general, each of which have four-year mandates. Things are organized so that the big contributors such as Germany and France will get, on average, two top posts; others such as Britain and Spain will get one each; and low contributors will only occasionally get a top position. At the end of the day, a brilliant Portuguese astrophysicist with experience — and proven skill — in managing big space-science projects is unlikely to be considered for the coveted post of director of science if the director of resources management is also Portuguese.

It's a similar situation at the European Commission, whose current constitution requires that each of the 27 member states has its own commissioner. There is no legal quota for positions below this level, but in practice a sharp eye is kept on the broad distribution of jobs, to avoid political problems developing. It is accepted that priority

for the higher positions that become vacant will be given to people from the new member states such as Bulgaria and Romania, until a reasonable geographic distribution is re-established. But at least all commission staff must pass a tough series of exams, which ensures some general level of competence.

Horse trading of positions in European organizations is a tough business, and is usually conducted by representatives of national governments who do not necessarily have the interests of the organization in question at heart. One of the most striking examples of this is the 2003 decision of the state council of the European Patent Office to split the six-year presidency between two heavily promoted candidates — one from France, the other from Britain — because they couldn't decide which august nation to offend. With the European Patent Office struggling to adapt to a massively increased workload and to new information technologies, it was highly inappropriate to weaken its leadership in this way.

Those European organizations that do fundamental research are, fortunately, less touched by the political interference than some of their counterparts. Even so, the nationality of the director of CERN, the European particle-physics laboratory near Geneva, is taken very seriously by its member states. Heads of its scientific divisions are selected entirely on merit, however. The European Southern Observatory seems to be free from national bias: four of its seven director-generals have been Dutch.

The geographical carve-up of key European positions is unfortunately intrinsic to multigovernmental organizations. Scientists should be aware of this political reality and should pressure their governments to ensure, as a minimum, that only suitably qualified candidates are put forward for significant posts. That at least will ensure that those with the 'right' nationality are unlikely to be completely wrong for a particular job that is important to science. ■

"Too often, key appointments are made on the basis of nationality rather than merit."

Safety clause

US research on bioweapons has expanded rapidly, without sufficiently transparent regulation.

In the aftermath of the September 2001 terrorist attacks, the US federal government wasted no time in allocating large amounts of resources to build facilities for research into dangerous pathogens that might be used by terrorists as bioweapons. But it is now emerging that there are problems with the way some of these facilities operate, which suggests that the overall process has been poorly

managed. These issues must be addressed before any further expansion goes ahead.

As construction of the first facilities got under way, some specialists were warning — at least privately — that many of the labs would lack people properly experienced in handling 'select agents', as microorganisms that could be used in bioweapons are now euphemistically called. Critics said that many of the institutions selected to host the labs lacked the capacity to manage the task.

Two weeks ago, these predictions seemed to come true. On 30 June, the Centers for Disease Control and Prevention (CDC) in Atlanta, Georgia, suspended research with select agents in five labs at Texas A&M University in College Station. The university is home to a

National Center for Foreign Animal and Zoonotic Disease Defense, one of several such facilities established by the US Department of Homeland Security.

The shutdown was the result of Texas A&M's failure to report in a timely manner that four members of staff had been infected by bacteria in early 2006. One worker fell ill from *Brucella* in February last year and two months later routine blood tests results revealed that three other staff members had been exposed to *Coxiella burnetii*. Both bacteria come from cattle; both are fever-causing pathogens and could be used in bioweapons.

Fortunately, no one was seriously injured, nor does any disease seem to have been spread to the wider population. Texas A&M insists that it is now improving its reporting procedures. And the federal government is treating the incident as it has others: a bump on the road to better biodefence through more laboratory research.

But a closer look at the case raises serious concerns. Texas A&M only got around to reporting the four exposures to the CDC on 9 April this year. This was just one month before officials from the Department of Homeland Security were due to arrive on the campus to discuss a major new biodefence contract: the \$450-million National Bio and Agro-Defense Facility, for which Texas A&M and its partner institutions are competing. The Department of Homeland Security, which will fund the complex, will this week narrow the 18 bidders down to a short list of three to five, with a final decision due in 2009. The new facility will replace the ageing Plum Island Animal Disease Center, which has operated for half a century off Long Island in New York state.

But even when Texas A&M made its report, neither the university nor the federal agencies found it necessarily to share news of the infraction with the general population. It was only on 26 June that the problem became public, when the Sunshine Project, a small watchdog group based in Austin, Texas, revealed details of the infections at Texas A&M. A week later, the watchdog disclosed associated problems at nine other laboratories nationally.

Back in April 2006, an audit from the Office of Inspector General at the US Department of Health and Human Services reported that 11 of the 15 biodefence labs it funds at universities had identifiable deficiencies in training, security and, most disturbingly, accountability. In fact, eight of the labs had accountability issues, including lax inventories for pathogens and inadequate controls on who can enter the laboratory.

If biodefence labs are to be run safely and successfully, the regulatory process needs to instil public confidence. The federal government should not have to be prompted by activists into telling the American public the truth about its workings. Many communities are inherently suspicious of these facilities and their mode of operation should be made as transparent as is realistically possible. The need to keep some of the technical details of the work secret should not be used to cover up salient facts about management and operations that ought to be squarely in the public domain. ■

"The federal government should not have to be prompted by activists into telling the American public the truth."

Hard to swallow

Is it possible to gauge the true potential of traditional Chinese medicine?

Researchers, practitioners and drug companies around the world are engaged in a complex, tentative dance over the best way to tap into the unknown potential of traditional Chinese medicine. The scientific community and the drug industry both tend to be sniffy about 'traditional' cures; yet there is a strong sense that millennia of practice in China — much of it barely documented — is likely to have yielded at least some treatments that work.

Pharmaceutical companies are understandably eager to enter a Chinese domestic market that was estimated by the Boston Consulting Group to be worth US\$13 billion last year, and growing fast. But they are tantalized by one opportunity above all: the prospect that the nation's traditional medicine might contain a number of potentially profitable compounds hidden somewhere in its arcane array of potions and herbal mixtures.

The task of finding these elusive gems has been approached in a typically reductionist manner, with researchers seeking single compounds that might have a role in treating specific diseases. Sometimes this has been successful: artemisinin, for example, which is currently the most effective treatment for malaria, was fished out of a herbal treatment for fevers. But such success stories have been few and far between.

So if traditional Chinese medicine is so great, why hasn't the qualitative study of its outcomes opened the door to a flood of cures? The most obvious answer is that it actually has little to offer: it is largely just pseudoscience, with no rational mechanism of action for most of its therapies. Advocates respond by claiming that researchers are missing aspects of the art, notably the interactions between different ingredients in traditional therapies.

Nevertheless, the drug industry is not exactly awash with promising new medicines at the moment. Perhaps as a result, the global regulatory process has become increasingly receptive to traditional approaches. In 2004, for example, the US Food and Drug Administration issued new guidelines on botanical drugs that made it much easier to get extracts into clinical trials if there was some history of prior use, and that obviated the need to characterize all compounds in an extract.

Some researchers in China and elsewhere, meanwhile, are advocating systems biology — the study of the interactions between proteins, genes, metabolites and components of cells or organisms — as a way to assess the usefulness of traditional medicines (see page 126). Constructive approaches to divining the potential usefulness of traditional therapies are to be welcomed. But it seems problematic to apply a brand new technique, largely untested in the clinic, to test the veracity of traditional Chinese medicine, when the field is so fraught with pseudoscience. In the meantime, claims made on behalf of an uncharted body of knowledge should be treated with the customary scepticism that is the bedrock of both science and medicine. ■

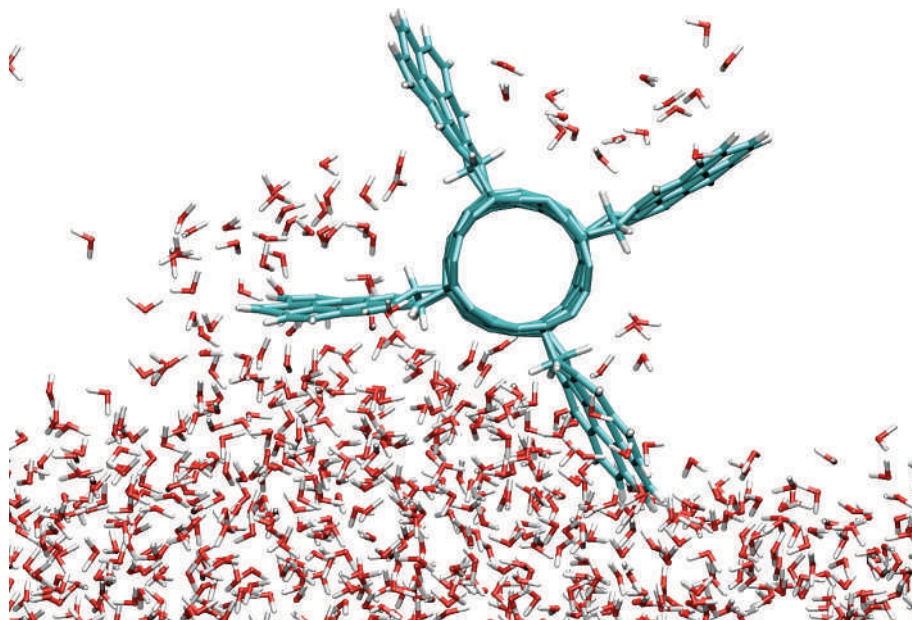
RESEARCH HIGHLIGHTS

Wheels in motion*Phys. Rev. Lett.* **98**, 266102 (2007)

Could a molecular-scale paddlewheel drive nanoscale liquid flows? Boyang Wang and Petr Král of the University of Illinois, Chicago, tested the idea using computer simulations.

The researchers modelled a paddlewheel made from a single-walled carbon nanotube with four flat molecular propellers attached around its circumference. As the tube spins, the paddles dip into a liquid layer below (pictured).

It was conceivable that the paddles might have simply disturbed the liquid molecules without forcing them forward, but the researchers say that the motion can indeed induce flow. The pumping rate depends on how the propellers interact with the liquid. For example, water sticks to hydrophilic paddles, clogging up the flow, whereas liquid long-chain alkanes create problems because the paddles struggle to drag them out of their entanglements.



AM. PHYS. SOC.

NEUROSCIENCE**Distracted by pain***Neuron* **55**, 157–167 (2007)

You might try doing some work to distract yourself from a toothache, but the pain will stop you concentrating. Neuroscientists have now collected the first neurobiological clues to the way pain comes to override the brain's cognitive resources.

Ulrike Bingel and her colleagues at the University Medical Center Hamburg-Eppendorf in Germany took images of the brains of 16 people who had been asked to identify objects in a series of clear or blurred pictures. During the task, some subjects were pricked on the hand and others were given memory exercises to do.

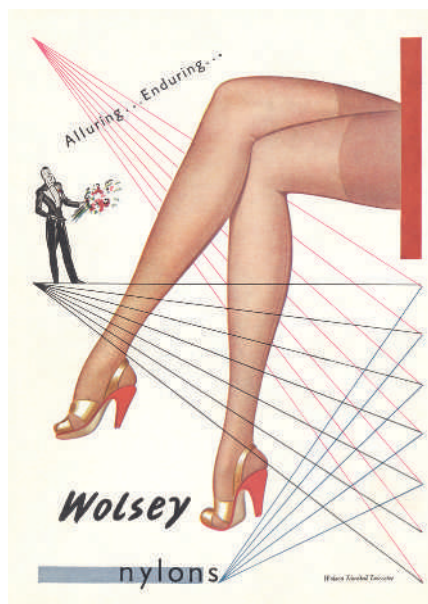
Both types of interference lowered activity in the brain area known as the lateral occipital complex, which is involved in processing visual data. This reduced the accuracy of subjects' object recognition. The researchers found that the brain regions directing this modulation differed for pain and working memory.

CLIMATE SCIENCE**Kept under ice***Science* **317**, 111–114 (2007)

DNA has been extracted from a silty layer at the base of an ice core drilled in southern Greenland, which reveals details of the pine forests and insect life the island once enjoyed. Found 2 kilometres down, the DNA is dated from between 450,000 and 800,000 years ago.

Greenland is known to have once been green — fossil plants have been found in the far north of the country dating to 2.4 million years ago. But the DNA provides the first evidence of what the landscape was like in regions now covered by ice.

The researchers, led by Eske Willerslev of Copenhagen University in Denmark, conclude that the absence of DNA younger than 450,000 years old means that this portion of land has been covered by ice ever since. This contradicts most interpretations of the sea-level record, which suggest that most of the Greenland ice cap melted during the last interglacial period.

**CHEMISTRY****Nylon from biomass***Chem. Commun.* doi:10.1039/b705782b (2007)

Researchers in the Netherlands have developed a distillation process that efficiently turns a chemical derived from the cellulose-rich parts of plants into a feedstock from which nylon polymers can be made.

Jean-Paul Lange of Shell Global Solutions in Amsterdam and his colleagues took γ -valerolactone, made from the biomass derivative levulinic acid, and mixed it with methanol and a strong acid catalyst to form the nylon precursor methyl pentenoate.

The boiling points of γ -valerolactone and methyl pentenoate differ by 80 °C. This makes the product easy to distil and collect. Previous routes to methyl pentenoate from biomass involved less efficient gas-phase reactions.

NANOTECHNOLOGY**Shape shifters***Nature Mater.* doi:10.1038/nmat1957 (2007)

Many of the potential uses of nanoparticles, for example as catalysts or in magnetic memories, demand that the particles have well-defined sizes and shapes. But that's not easy to arrange at such small scales, and it has previously been largely a matter of trial and error.

Peidong Yang of the University of California, Berkeley, and his co-workers have now shown how to control the size and shape of metal nanocrystals.

When they deposit palladium on to

BRIDGEMAN ART LIBRARY

nanocubes of platinum, each cube acts as a seed for uniformly sized cubic palladium nanoparticles. Adding nitrogen dioxide to the mixture changes the particle growth rates in different crystallographic directions to give cuboctahedral or octahedral nanocrystals instead.

MATERIALS SCIENCE

Film conductor

Nano Lett. doi:10.1021/nl070477 (2007)

Science's newest wonder material, graphene, can be used to make glass conduct electricity, say researchers led by Rod Ruoff at Northwestern University in Evanston, Illinois. Their transparent and conductive composite material containing graphene — sheets of carbon one atom thick — could find applications in devices such as solar cells.

Transparent materials can be given a conductive coating of indium tin oxide, but indium is expensive. Previous carbon-based coatings have had problems with adhesion and opacity.

Ruoff's group combined graphene-oxide sheets with silica to form a solution that could be spin-coated quickly and cheaply on to glass.

GENETICS

Anemone genome

Science **317**, 86–94 (2007)

The genome of the starlet sea anemone, *Nematostella vectensis* (pictured below), has helped researchers to tease out details of early animal evolution.

Daniel Rokhsar at the US Department of Energy Joint Genome Institute in Walnut Creek, California, and his collaborators have put together most of the 450-million base-pair genome, revealing some 18,000 genes.

By comparing this sequence with those of distant relatives, they put together a genomic picture of the last common ancestor of cnidarians (such as anemones and jellyfish) and bilaterians (such as flies, worms and humans). Humans and the starlet sea anemone seem to have kept some features of this putative primordial genome that have been discarded by flies and worms.



MEDICAL RESEARCH

Drugs doubled up

Cancer Cell **12**, 81–93 (2007)

Researchers have found a way to prolong the effectiveness of targeted cancer therapies in a mouse model of lung cancer.

The therapies gefitinib and erlotinib inhibit a protein called epidermal growth factor receptor kinase, thus shutting down tumour growth in some non-small-cell lung cancers. But the tumours soon acquire secondary mutations that make them resistant to the drugs.

Geoffrey Shapiro and Kwok-Kin Wong of the Dana-Farber Cancer Institute in Boston, Massachusetts, and their co-workers show that a drug designed to combat these secondary mutations, HKI-272, doesn't work well on its own in mice engineered to get cancer. But when combined with an existing drug, rapamycin, HKI-272 quickly fights back tumours, suggesting that this one-two punch may serve as a back-up to extend the benefits of targeted therapies.

MICROBIOLOGY

Birth of a plague

Cell. Microbiol. doi:10.1111/j.1462-5822.2007.00986.x (2007)

Scientists may have identified a step in the evolutionary path that produced the bubonic plague.

Bubonic plague is caused by the bacterium *Yersinia pestis*, which is transmitted to humans by rat fleas. B. Joseph Hinnebusch of Rocky Mountain Laboratories in Hamilton, Montana, and his colleagues report that a related bacterium, *Yersinia pseudotuberculosis*, is toxic to fleas, causing diarrhoea.

They think this toxicity is due to a protein that acts in the flea gut. The researchers ruled out Tc toxins, a class of bacterial toxins with known insecticidal activity, as being to blame, leaving the identity of the protein unknown.

The group suggests that modification of the toxic protein may have enhanced flea-borne transmission and played a part in the genetic divergence of the two bacteria.

JOURNAL CLUB

Michael A. Marletta
University of California,
Berkeley, USA

A biochemist marvels at a molecule that shares his love of playing with fire.

I like to capture my students' attention by recounting how my early fascination with fire inspired my interest in the stability of sugars.

Glucose will 'burn' to carbon dioxide and water, liberating lots of energy. But it is stable enough that you can stamp on it without triggering the reaction — the energy barrier to the reaction is too high.

In my research, I am interested in how biology harnesses and controls oxygen reactivity. Most reactions, such as burning glucose, are held back by an energy barrier to getting things started. Enzymes can bypass this, finding a lower energy route through some reaction intermediate, to carry out a 'controlled burn'. Their control is not perfect, sometimes causing damage to both themselves and surrounding molecules, but by and large it works.

Typically, these enzymes have metal or organic components, which drive the oxidation. I often tell students that enzymes need their metal and organic cofactors because the 20 naturally occurring amino acids cannot carry out all the chemistry. Two recent papers shake that belief.

The surprise comes from the enzyme DpgC, which is involved in the biosynthesis of the antibiotic vancomycin. The first paper (C. C. Tseng *et al. Chem. Biol.* **11**, 1195–1203; 2004) reports that DpgC uses oxygen in a complex dioxygenase reaction with no bound metal or organic cofactor.

More recently, researchers reported the structure of DpgC and confirmed that it has no cofactor (P. F. Widboom *et al. Nature* **447**, 342–345; 2007). They find that the enzyme has a structure known as an oxyanion hole, which helps to stabilize the reaction intermediate.

I am still amazed that DpgC does oxygen chemistry with no help — and my students should be too.

Discuss these papers at
<http://blogs.nature.com/nature/journalclub>

NEWS



I. VAN MIDDELKOOP/CAPRISA

Research at this South African clinic includes a controversial microbicide-gel trial.

HIV trial doomed by design, say critics

Controversy over a ground-breaking study of an experimental HIV prevention tool has underscored the field's need to revamp its approach to clinical trials.

On 23 May, researchers at the Centre for the AIDS Programme of Research in South Africa (CAPRISA) began a US\$13.5-million study to test whether a microbicide gel containing the antiretroviral drug tenofovir prevents women from becoming infected with HIV during sex. The trial, funded by the US Agency for International Development (USAID), is the first to test a so-called second-generation microbicide — one that specifically targets the AIDS virus.

Researchers and advocates say that microbicides in the form of gels or creams, applied to the vagina, could provide crucial protection for women who cannot negotiate condom use with their partners. Currently, 11 microbicide candidates are being tested in clinical trials. But some experts are concerned that the "CAPRISA 004" trial, which includes 980 women at two sites in South Africa, is doomed because of its design.

"The microbicide development field cannot afford to take a further hit," says virologist Mark Wainberg of McGill University in Montreal, Canada.

That is because three 'first-generation'

microbicides, which make the vagina inhospitable to a range of microbes, have already failed efficacy trials, and two — cellulose sulphate and nonoxonyl-9 — actually increased women's risk of infection (see *Nature* **446**, 12; 2007). Investigators are due to release the final analysis of the cellulose sulphate trial at the International AIDS Society meeting in Sydney, Australia, later this month. But scientists say the failure of that trial highlights the need to proceed cautiously with further microbicide studies.

The most controversial issue in the CAPRISA trial is the dosage schedule. This requires women to apply the gel once within 12 hours before sex and again within 12 hours after sex. The schedule is designed for women whose partners are home for short periods of time, says study leader Salim Abdool Karim of the University of KwaZulu-Natal in South Africa.

But opponents say there isn't enough published evidence from trials in animals or humans to support this dosage schedule. Some would prefer the women to apply daily doses of the gel. Karim disagrees, saying that "getting these women to use a product every day is going to be a challenge."

Critics say the proposed dosage regime will result in a poor trial outcome. "It will be hard to link the data to the way the gel is being used, which will potentially make it very difficult to make any interpretation of the data," says microbicide researcher Robin Shattock at St George's, University of London.

The concern has led scientists and others, including Renee Ridzon of the Bill & Melinda Gates Foundation in Seattle, Washington,

and Zeda Rosenberg of the International Partnership for Microbicides based in Silver Spring, Maryland, to call for the CAPRISA investigators to rethink their approach. Such concerns prompted the US

Global AIDS Coordinator to convene a telephone conference on 6 June with scientists, advocates and US government agencies to discuss the trial design — a first for a USAID-funded microbicide study.

After the telephone conference, the Office of the Global AIDS Coordinator declined to modify or stop the trial. But scientists are still urging the CAPRISA investigators to reconsider, and Karim has agreed to discuss the issue further, at least informally with some of the critics. "We

"The microbicide development field cannot afford to take a further hit."



WHALING MADE PENGUINS EAT KRILL
Birds changed diet after humans killed shrimp eaters.
www.nature.com/news

S. EMLIE

have put the issue on the table and I feel we are on our way to resolution," he says.

The situation has caused leaders in the microbicide field to renew calls for better ways of reviewing and coordinating trial plans. Scientists in the field already share and discuss ideas and data through formal and informal meetings, but there is no mechanism for reviewing and disseminating plans for clinical trials. This means that institutions sometimes run redundant trials that are not always backed by good evidence, says immunologist John Moore of the Weill Cornell Medical College in New York.

"Historically, the microbicide field has gotten it wrong by doing multiple simultaneous trials of similar products, and the CAPRISA episode represents a boiling-up of frustration in the field," Moore says.

The study had already been approved by many bodies — including the South African Medicines Control Council, and ethics review boards convened by the University of KwaZulu-Natal and by Family Health International, the group that is running the trial. USAID staff decided to grant funding for the study in April after conducting its own internal review, says Jeff Spieler, senior science adviser in USAID's office of global and reproductive health.

But few preclinical results have been made publicly available, leaving scientists and advocates uneasy about the trial. Normally, the development of a product such as a microbicide would be sponsored by a private company, which would not be obligated to consult the public beyond regulatory agencies. But such companies are not interested in developing microbicides because they are aimed at poor people, so the field is funded largely by limited public resources. Advocates argue that this means the public deserves full access to all the relevant data.

"We need a transparent, accessible and international mechanism of peer review and reflection for making decisions that are of fundamental importance to the microbicide field," says Lori Heise, director of the Global Campaign for Microbicides, an advocacy group based in Washington DC.

Spieler says USAID has decided to consult external reviewers for future trials. And the National Institutes of Health's Office of AIDS Research is setting up a microbicide working group. But it is not clear whether this will include a broad enough range of international expertise to ensure that good decisions are made, Heise says. And others say the field is already facing a crucial turning point.

"The microbicide field is drinking in the last-chance saloon," Moore says. "If it has many more problems, it's finished."

Erika Check

Conserving life and livelihood

Last week, more than 1,600 people involved in conserving Earth's flora and fauna came to Port Elizabeth, South Africa. It was the first time that the Society for Conservation Biology, based in Arlington, Virginia, had met in Africa, and the setting raised a challenging question: does conservation help poor people?

Some researchers think that conservation work will naturally and inevitably benefit local people, because it will sustain resources over time. But conserving animals and plants often means restricting access to them. And as Yaa Ntiama-Baidu, a Ghanaian conservationist, asked those gathered at the opening plenary session: "Do your conservation projects make a difference in village life in Africa?"

To try to get some data with which to answer her own question, Ntiama-Baidu, who works with the conservation group WWF and the Ghanaian government, looked at 50 projects in Africa. Of the project managers surveyed, 92% thought that they were making a difference on the community level. But projects that tried to measure the effects — such as a wetland conservation project in Ghana that measures the number of

new enterprises created by an eco-tourism effort — were few and far between. Hardly any of the 50 projects had any built-in way to quantify or demonstrate their benefits. "Why is it that we do not have concrete data to support this?" Ntiama-Baidu asked.

In a session on the link between conservation and poverty, Peter Kareiva of Seattle, Washington, chief scientist at the Nature Conservancy, presented an analysis of almost 200 development projects run by the World Bank, some of which had environmental components. Encouragingly, he found that development projects with built-in conservation goals were no less effective than those without them, as measured by the World Bank's evaluations. But his data do not address the issue of how effective conservation projects are if they include specific development goals in their remit.

Many hope that local researchers will take the lead on conservation projects. "If Africans have the empowerment and the tools, they will have to make these decisions themselves," says Jonathan Adams, also of the Nature Conservancy.

In South Africa, there

are hopes that more black people will take careers as conservation scientists. According to Brian Huntley, an environmental adviser to the South African government, black conservation scientists are only now starting to emerge, more than a dozen years after the end of apartheid, and he thinks that their number will increase exponentially.

But for many South Africans, continued poverty and the lingering social effects of apartheid are daunting obstacles to such a career. "When we were kids, we weren't allowed to go to the aquarium or anything like that, so how were we to learn that saving nature was important?" asks Mncedi Nkosi, a young, black, freshwater ecologist at Ezemvelo KZN Wildlife, a province-level public-private conservation group. "Most people still are more worried about socioeconomic issues, and they don't really understand what I do. I sometimes just say that I clean water for a living."

About one-third of the 700 papers and posters at the conference were presented by Africans, according to the meeting's organizer, Graham Kerley.

Emma Marris

A. JOE/AFP/GETTY



Few development projects measure the effects they are having on village life.

SCORECARD

**McDonald's**

The fast-food giant has vowed to convert all its UK distribution trucks to run on biodiesel — created from its huge stock of cooking oil.

**Albert Gore**

Al Gore's son has been caught speeding at 160 kilometres per hour in a Toyota Prius hybrid car. The offence doubles as a handy demonstration of low-carbon technology in the week of his Dad's Live Earth concerts...

NUMBER CRUNCH

3 crew members are currently on board the International Space Station.

6 will be the size of the crew by 2009.

\$19 million is the amount NASA has just spent on a new toilet system to keep them in the lap of luxury.

SHOWBIZ NEWS

The Horrible Girls from CERN

Fans of particle-physics-based pop music rejoice: later this month, CERN — Europe's particle-physics laboratory near Geneva, Switzerland — will host its annual 'Hardronic Festival', featuring the long-awaited return of the band Les Horribles Cernettes. Alongside hits such as *Mr Higgs* and *Every Proton of You*, their photo is rumoured to be the first image ever posted on the Internet.



ON THE RECORD

"Another female. It wasn't me."

The response of Michigan state forensic scientist Ann Chamberlain-Gordon during her divorce trial, when asked what she had found when she used resources from her lab to analyse the DNA in her husband's underwear.

Sources: BBC, Toronto Star, Associated Press, motls.blogspot.com

Austria's science institute gathers steam

Austria will, after all, get the élite graduate-level institute it wants — even though plans seemed on the verge of collapse last year.

On 12 July, an 11-member scientific advisory board was expected to approve the focus for the planned Institute of Science and Technology. All research themes will be interdisciplinary, with focuses in the areas of biology, medicine and the physical sciences. Detailed research programmes will be defined by the institute's scientists, who are now being headhunted. The overall organizational structure follows the example of the Weizmann Institute of Science in Rehovot, Israel.

That's not surprising, as it is Haim Harari, a theoretical physicist and the former president of Weizmann, who has been driving plans for the institute. Originally spearheaded by Anton Zeilinger, head of physics at the University of Vienna, the project lost steam last year when Zeilinger left, complaining about the influence of local politics. Of particular concern was the decision to place the institute in the town of Gugging, 45 minutes from Vienna, at the site of a former psychiatric hospital.

Harari, who joined the project in late 2005, insisted that politicians should give the scientific community more control over the project. And although the institute will still be built at Gugging, fears have abated that the distance may make it less attractive as a workplace. Harari says that the campus "has fantastic potential — when you actually go there you can see that". And the institute may "come to life in a cloistered atmosphere," suggests physicist Olaf Kübler, chair of the scientific

advisory board and former president of the Swiss Federal Institute of Technology — ETH Zurich.

Zeilinger has since returned to the scientific board and board of trustees, although he refuses to comment on the political decisions about the location.

With the decision to proceed now firmly in place, outside observers also see the need to gather leading scientists together. "Now it is necessary for the Institute of Science and Technology to reach a chain-reaction mass, so that it can get on with its work in the absence of closely neighbouring institutes," says Christoph Kratky, a molecular biologist from the University of Graz and head of the Austrian science fund FWF in Vienna.

The institute will not be part of the existing group of Austrian universities, which is largely seen as overly bureaucratic and not able to offer competitive packages to top scientists.

Germany and Italy, whose universities share the same historical system, have also toyed with the idea of building an elite graduate school from scratch. Germany, however, abandoned the idea in favour of topping up funding of existing universities that perform well, and the Italian Institute of Technology is floundering.

Government and industrial sources are supporting the Austrian project to the tune of 500 million euros over 10 years. If all goes to plan, the institute will eventually employ 500 people. With construction beginning in autumn, scientists who don't need complex equipment are expected to start work next year.

Carina Lenotti



Haim Harari insists on scientific involvement.

The Institute of Science and Technology will have an idyllic setting, replacing an existing psychiatric hospital.



INST. SCI. TECH. AUSTRIA



CANDIDATE DRUG FOR BIPOLAR DISORDER

A designed alternative to lithium shows early promise.

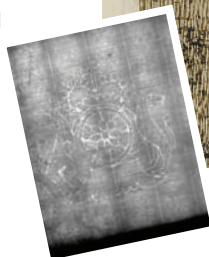
www.nature.com/news

SNAPSHOT Marks of distinction

During the eighteenth century, Giovanni Domenico Tiepolo (son of the more famous Giovanni Battista Tiepolo) was commissioned by the prince-bishop of Würzburg to depict the biblical story of the flight of the Holy Family from King Herod. But did he start the series of 24 etchings, called *The Flight into Egypt*, in his hometown of Venice, Italy, or only after he arrived in Würzburg, Germany, at around 1750?

Art historians in Freiburg, Germany, have now acquired the technology to help them solve this long-standing mystery. The method allows watermarks to be seen *in situ* without damaging the paper in which they are embedded. It was developed by scientists at the Technical University Braunschweig and the Fraunhofer Institute for Wood Research, also in Braunschweig, and was first tested last year on Rembrandt sketches.

Paper mills have characteristic watermarks that are visible when the paper is held to the light, although not when they are



obscured by ink. X-ray analysis can be used to identify obscured

watermarks, but it needs special conditions to protect the paper from radiation. Delicate materials therefore usually need to be transported, and museum curators are often reluctant to do this.

In the new method, a plate warmed to 35–40 °C is placed

behind the paper for one second, during which an infrared camera captures the heat passing through it. The outline of the watermark is revealed because it lets more heat through than does the rest of the paper. The picture here illustrates an early etching for the series, together with its watermark (inset), as exposed by the thermography technique.

It is part of the exhibition Giandomenico Tiepolo: *The Flight into Egypt*, which opens this week and runs until 16 September at the Augustiner Museum in Freiburg. All the watermarks stem from a paper mill not far from Würzburg, so Domenico almost certainly started work on the series after he arrived in Germany.

Alison Abbott

H.-P. VIESER/AUGUSTINERMUSEUM

French universities to gain control

"In the race against Stanford, Cambridge or Harvard, French universities run with their laces tied together and a backpack full of stones." So said Nicolas Sarkozy in the run-up to the French presidential election, as he pledged to reform the country's archaic university system. As the new president, Sarkozy has now personally weighed in on a reform bill that will be fast-tracked through parliament this summer.

The bill, adopted by the cabinet on 4 July, is historic as it would make France's 85 public universities much more independent, largely freeing them from the current centralized state control.

Sarkozy has also confirmed that universities will receive an extra €5 billion (US\$ 6.8 billion) over the next five years. Most people agree that this sum, and much more, is badly needed. Whereas the elite *Grandes Ecoles* — which

scoop the best few per cent of students — are well-heeled, the underfunded universities must cope with most of the remainder.

The bill would allow universities to own and manage their own buildings, to control their budgets, and to hire and set salaries as they see fit, all of which are currently controlled by the science and higher-education ministry. At present, a star biologist might earn no more than a philosopher of the same bureaucratic grade. Top international research talent often passes French universities by.

The bill would also modernize governance. University presidents have had few real powers, and whereas Anglo-Saxon universities typically form committees to headhunt the best leader, in France the presidents are elected by

130 members of the various university panels, and can serve only one four-year term. The university's direction is decided largely by a 60-strong board of directors that is often highly politicised.

Under the new law, the board would be streamlined to 20–30 people, with an absolute majority agreeing on a president. The president, elected for a maximum of two four-year terms, would have executive power over almost all university affairs.

The bill is just the start of broader university reform, says prime minister François Fillon, who has described the future of French universities as the most important item on his domestic agenda.

Declan Butler

"French universities run with their laces tied together and a backpack full of stones."



**TB DIAGNOSIS CHANGE
CAUSES CONFUSION**
Identifying drug-resistant
tuberculosis is a global
problem.
www.nature.com/news

Fly library boosts gene tool supply

Researchers have developed a library of transgenic fruitflies that will allow them to investigate genes in specific tissues or developmental stages across the whole of the insect's genome. The library, which consists of genes known as RNA interference (RNAi) transgenes, is described this week in *Nature* (see page 151). The transgenes act to silence their target genes, enabling researchers to work out the function of genes that would be lethal if switched off throughout the whole organism.

The fruitfly *Drosophila* is a model organism for developmental biologists. And *Drosophila* researchers have been using RNAi for years, but typically for testing only a few genes at a time. "We didn't invent anything here," says Barry Dickson, a molecular biologist at the Research Institute of Molecular Pathology in Vienna and lead author of the latest paper. "We were just crazy enough to go for the whole genome." Previously, genome-wide screens were used only in cell-culture experiments, and results had to be confirmed in live flies.

Now, Dickson and his colleagues have created a collection of 22,270 lines of transgenic fruitfly in which 12,088 genes can be silenced — about 88% of the fly's predicted protein-coding genes. Each fly contains an RNAi transgene that silences a targeted gene only in the presence of a protein called GAL4. Crossing the RNAi-containing strain with a strain that expresses GAL4 in a specific tissue activates RNAi only in that tissue. In *Drosophila*, unlike in many other insects, the RNAi signal does not travel from cell to cell.

The lines offer researchers the chance to trawl for tissue-specific functions of genes involved in fruitfly development. Dickson's lab, for example, studies a gene called *fruitless* that regulates mating behaviour. The gene's role went undiscovered for years because flies lacking it don't survive. But the auspicious discovery of fly mutants in which one form of the *fruitless* protein was absent in a select group of neurons revealed that the gene was crucial in triggering courtship behaviour in male flies.

In their latest study, Dickson's team took a sample of 50 genes that were lethal when silenced in all the fly's cells, and silenced them selectively in the eye, wing or thorax. Only a third of the silenced genes were still lethal. Another third were found to cause specific developmental defects in the tissues in which they were expressed. Researchers have already started to use the lines to search for genes

**"We were just crazy
enough to go for the
whole genome."**



Tiny triumph: a collection of 22,270 lines of fruitfly are available for the study of developmental genes.

involved in eye development, courtship behaviour and the RNAi mechanism itself.

"This library is a great, great tool," says Michael Boutros, a geneticist at the German Cancer Research Center in Heidelberg. "I think it will be heavily used in the future." Boutros's lab plans to use the strains to study the *Drosophila* immune system.

The library generated by Dickson's group is available through the Vienna *Drosophila* RNAi Center, which opened in April. It already sends out about 2,000 strains each week. Potential users would do well to order soon, though: because the centre doesn't have the funds to maintain duplicate lines, there is no back-up when a vial of flies dies. The facility is currently losing RNAi strains at the rate of about two per week.

That's a common problem for the *Drosophila* community, says Claude Desplan, a molecular geneticist at New York University. "Quite a number of people are coming out with large collections of flies," he says. "But people don't know how they're going to maintain them." Dickson says he has been unable to secure government funding for the stock centre and must rely on user fees. The going rate is about €10 (US\$14) a vial, each containing tens to hundreds of flies.

Other *Drosophila* RNAi libraries are also in the works. One, created by Ryu Ueda of the National Institute of Genetics in Shizuoka,

Japan, contains 13,000 strains so far. "This is useful because the two projects together give us more strength to cover the full genome," says Christophe Antoniewski of the Pasteur Institute in Paris. Antoniewski has been using Ueda's lines to study how RNA regulates gene expression, and plans to order lines from Dickson as well.

A third library, still in early development, is being created at Norbert Perrimon's lab at Harvard Medical School in Boston. Whereas Dickson and Ueda created their libraries by randomly inserting the RNAi transgenes, Perrimon aims to target his transgenes to a specific region in the genome. This ensures that a transgene will not land in a region of the genome that may affect the gene's expression. "The level of expression is critical," says Perrimon, "and expression is very position-specific." RNAi transgenes that aren't fully expressed can lead to false-negative results.

Meanwhile, Dickson's group has found that low expression of an RNAi transgene can often be counteracted by overexpressing a gene called *dicer-2* that is involved in the RNAi pathway. "This usually helps, but it can also be dangerous," says Dickson. Overexpressing *dicer-2* can increase the risk of inadvertently silencing other genes.

Such caveats are always present with high-throughput procedures, says Desplan. "Until we see how people use it, there's a wait and see position," he says. "But I'm very optimistic." ■

Heidi Ledford

C. NURIDSANY & M. PERENNIOU/SPL

Activated eggs offer route to stem cells

A technique that could sidestep many of the limitations and ethical concerns that plague the production of human embryonic stem cells was unveiled late last month by a team of US and Russian researchers. The group managed to derive embryonic stem cells from an unfertilized egg — avoiding the need to use viable embryos.

“This is one of the papers we’ve been waiting for,” says George Daley, a stem-cell expert at the Children’s Hospital in Boston, Massachusetts, who was not part of the collaboration. But he and others caution that more work needs to be done before the method can be tried in the clinic. And the technique’s dependence on access to fresh, unfertilized eggs — a controversial resource — is also likely to limit its application.

The team, led by Elena Revazova and Jeffrey Janus of Lifeline Cell Technology, a biotechnology company in Walkersville, Maryland, created embryos by activating an unfertilized egg using chemicals rather than fertilizing it with sperm (E. S. Revazova *et al.* *Cloning Stem Cells* 9, doi:10.1089/clo.2007.0033; 2007). Some lizards, birds and insects can reproduce without the need for a male — a process known as parthenogenesis — but mammalian parthenotes cannot develop past an early-stage embryo. Revazova and her team managed to get the embryos to develop far enough for them to yield stem cells.

The resulting cells express the same proteins as normal embryonic stem cells and proliferate for an extended period of at least 10 months, the team says. The cells also seem capable of developing into the three primary tissue types. “For all intents and purposes, they are the same” as regular embryonic stem cells, Janus says.

He hopes that the technique will help solve one of the problems associated with stem cells derived from embryos left over from *in vitro* fertilization (IVF) procedures. Stem cells derived from IVF embryos would not be genetically matched to patients and so if used to treat them would probably provoke an immune response. But, says Janus, stem cells from the parthenogenetic embryos should be compatible with the donor’s immune system. He admits, however, that “this will have to be proven in a clinical trial”.

Daley has already done a proof-of-principle study in mice, in which he transplanted parthenogenetic embryonic stem cells back into the mice without provoking an immune reaction (K. Kim *et al.* *Science* 315, 482–486; 2007).

But the cells would be a direct match only



M. STOIKOVIC/SPL

Who’s the daddy? Parthenogenesis could provide a more acceptable way to derive embryonic stem cells.

for women, of course, although they might be a good enough immunological match to work for relatives of the donor women too. An alternative approach might be to create a bank of parthenogenetic cells that represent all the immunological ‘families’ — although estimates for the number of cell lines this would require range from 50 to 1,000.

Revazova and her colleagues next plan to make cells that could treat liver disease, diabetes and the eye condition macular degeneration. But there are potential problems. Cells express genes differently depending on the parent who provided them. So the next step is to see whether the parthenogenetic cells behave the same as normal cells once they have been turned into a specific tissue type, says Jose Cibelli, a stem-cell researcher at Michigan State University in East Lansing.

Daley notes that there have been problems in mice. When parthenogenetic cells were mixed with normal embryonic cells to form a ‘chimaera’, abnormalities caused poor tissue growth. “It is unclear whether this will be a problem if fully formed tissues are made *in vitro* and transplanted into an adult,” he adds.

Janus says the team hopes to begin safety trials in animals in early 2008, using retinal cells produced by the technique. If all goes well, similar trials in people could begin in late 2008 for treating macular degeneration.

The researchers are optimistic about the efficiency of their technique, saying that they established 6 cell lines using 46 eggs from 5 women. “I am convinced the work can be repeated,” says Cibelli.

Use of the technology will be limited by its source: women who bear healthy eggs. The ethics of getting eggs will also be a problem. Janus says the team recruited donors in Russia from women who were undergoing IVF procedures. The team’s paper says no financial payment was made to the donors, but Janus says they did cover some costs of the IVF treatment.

Last year, a team at the University of Milan in Italy claimed to have produced parthenogenetic embryonic stem cells, but their results have yet to be published (see *Nature* 441, 1038; 2006). And significant steps were recently made for another promising method of producing patient-specific cells using genes inserted into mouse skin cells to reprogramme them to an embryonic state (see *Nature* 447, 618–619; 2007).

The method that would give cells that are the best match for the patient — cloning — has so far failed. With more than 2,000 eggs, South Korea’s Woo Suk Hwang failed to produce a single cloned cell line. Ironically, he produced parthenogenetic stem cells — but his team failed to recognize them as such and carried out none of the analyses that could have won the group kudos for the achievement.

David Cyranoski

“I am convinced the work can be repeated.”

NASA offers fresh lease of life to old spacecraft

NASA is planning to redeploy two craft that have already performed their initial missions. The craft will be used to hunt for rocky planets and to explore comets.

The space agency will give one of the craft, the comet probe Deep Impact, two new charges. Beginning in September, it will measure nearby stars as giant planets eclipse them, looking for signs of associated rings, moons and Earth-like planets. Then late next year, Deep Impact will fly by Boethin, a small, unexplored comet, to follow up on observations of other comets.

The second spacecraft, Stardust, which originally collected comet samples, will now head to the comet Tempel 1. It is expected to reach it in 2011, and measure the crater made when Deep Impact hurled a 370-kilogram probe into the comet in 2005. The resulting debris obscured measurement of the crater itself at the time.

The redeployments are expected to cost about \$30 million each, or about 10% as much as new missions would do. "You get great science for very little risk and almost no cost," says Joseph Veverka, the Cornell University astronomer leading the Stardust redeployment.

Zurich university appoints chemist as rector

The Swiss Federal Institute of Technology (ETH) in Zurich has completed recruitment for its top positions with the appointment of Heidi Wunderli-Allenspach, a chemistry professor there, as the

university's first-ever female rector.

She will work alongside the institute's new president, Ralph Eichler, who was appointed in May. Eichler, a physicist, is currently director of the Paul Scherrer Institute, a sister research institute. Both will take office on 1 September.

The ETH has lately been rocked by internal disputes that culminated in the resignation of its last president, Ernst Hafen, in November 2006, and formal complaints to the government about perceived unfairness in the allocation of federal university funds.

Wunderli-Allenspach says that this is now a time for consolidation and reconciliation — and, perhaps, for hiring more women. "It is not a hostile place here for women, but we



Incoming rector: Heidi Wunderli-Allenspach.

Survey uncovers troupe of endangered monkeys

Conservationists say that they have discovered the largest-known population of the grey-shanked douc, or *Pygathrix cinerea* (pictured), in central Vietnam.

A team from Conservation International and the conservation group WWF said on 2 July that its members had tracked at least 116 specimens of the endangered monkey species in three surveys of Que Phuoc Commune in Quang Nam — the largest number ever observed in one place. It is thought that there are fewer than 1,000 of these monkeys left.

According to conservationists, the existence of this troupe increases the species' chances of survival.



T. NADLER/ENDANGERED PRIMATE RESCUE CENTRE

have to make a bigger effort, earlier on, in recruitment," she says.

Google offers the Earth to help in public education

Google has launched an initiative to help not-for-profit and public-sector groups, including scientists, use its Google Earth platform for public education.

The Google Earth Outreach programme will provide free tools and tutorials to assist people with displaying complex data sets on Google Earth, and will also make Google's own developers available to help with selected projects.

Scientists will be an important target, says programme manager Rebecca Moore, adding that the goal is to "demystify" the computing needed to convert data for ready visualization. Researchers are already using Google Earth for some scientific applications, such as the modelling of volcanic ash plumes. But the outreach effort will focus on issues of broader public interest, such as global warming, Moore says.

India pledges major role in fusion-energy project

The Indian cabinet has formally approved the country's participation as a full member in the ITER international fusion-power project, at a cost of more than US\$625 million over ten years. India will supply parts such as cryostats, power supplies and beam diagnostics.

The government said it would set up a board, whose chair will be selected by the Institute of Plasma Research in Ahmedabad, to manage its participation, the cost of which amounts to about one-sixth of India's annual research budget. "Considering India's large energy needs in the future, participation

in ITER will allow India to leap forward in terms of our capability in fusion energy," a government statement said.

The United States, Russia, the European Union, Japan, China and South Korea are the other members of the project, which is being built at Cadarache, France.

US wetlands run risk of losing protected status

The Bush administration has introduced guidelines that environmentalist groups claim will strip large swaths of US wetlands of their protected status.

The rules, issued last month, sparked the latest skirmish in a long-running dispute between the administration and environmentalists over the 1972 Clean Water Act. The act provides protection for major rivers, but not necessarily for wetlands that feed into them via tributaries. The latest guidance instructs officials at the Environmental Protection Agency and the Army Corps of Engineers to protect only wetlands that are directly connected to major waterways.

"Most wetlands don't have a continuous surface connection," says Joan Mulhern, a senior lawyer with Earthjustice, a non-profit environmental-law firm based in Oakland, California. Wetlands should be protected irrespective of that, she adds.

Nature wins major award in Spain

Nature has won the prestigious Príncipe de Asturias Award, established by His Royal Highness the Prince of Asturias, heir to the throne of Spain.

The award is the best-known cultural prize in the Spanish-speaking community. The awards honour individuals, groups or institutions whose creative work or research represent a significant contribution to universal culture in the scientific, technical, cultural, social and humanistic fields. *Nature* shares the 2007 Award for Communication and Humanities with the journal *Science*.

Technical advice

Scientists tend to complain that Congress rarely pays heed to what they have to say. But the issues are often as much about values as they are science, says **David Goldston**.

How should Congress seek advice from scientists? This question has been mired in controversy ever since the Republicans abolished the Office of Technology Assessment (OTA) in 1995 when they took control of Congress. The office was set up by Congress in 1972 to write reports on technical issues, and the resulting documents were regarded highly by science-policy experts. The Republicans killed the OTA mainly to save money — it had been receiving about \$20 million annually in its final years — although some conservatives also complained that the agency dispensed biased advice.

The Democratic takeover of Congress in last year's elections revived the hopes of those pushing for the OTA's re-establishment. Last month, however, the Democrats decided that it would be too costly to include money for the agency in next year's appropriations. As an alternative, the spending bills include funding — \$2.5 million in the House bill and \$750,000 in the Senate version — for the Government Accountability Office (GAO), an investigative arm of Congress, to work on technology assessments. In recent years, the GAO has undertaken a number of technology studies at the request of Congress, but the results have received scant attention.

For many US scientists, the demise of the OTA has taken on inordinate symbolic significance. Scientists often suggest that in eliminating the agency, Congress chose a path of wilful ignorance that has led to poor decisions over the past decade. But has the absence of the OTA really hampered policy-making? Not particularly. Congress is awash with information provided by scientific groups, and it still formally seeks scientific guidance — particularly from the National Academies, which arguably have more prestige and credibility on Capitol Hill than the OTA had. And reports from the academies can make a difference. For example, the 2006 report on the palaeoclimate record, specifically on the 'hockey-stick' graph (see *Nature* **441**, 1032–1033; 2006), helped quieten congressional debate over whether recent decades have been unusually warm.

Other reports have been equally prominent, if less decisive. The academies' 2002 report on fuel-economy standards has become the bible on that subject, although, like the Bible, it is quoted by all sides. That's partly because of the



PARTY OF ONE

report itself — it concluded, for example, that the standards had cost lives in the past but that, because of new technology, they needn't in the future — but it is also because scientific information does not usually point ineluctably to a single conclusion on policy.

Policy-making needs to be informed by both science and values. Is stem-cell research ethical? That's not a science question, although one needs to understand the potential of stem-cell research to answer it. Should clean-air standards be strengthened? That is not a science question, but one needs to know what researchers think the health impacts of dirtier air would be.

Often, when scientists complain that Congress is ignoring science, what they really mean is that Congress is making policy calls they don't like. The notion that Congress would make different decisions if the OTA were around is another way of saying: "If you knew what I know, you'd think what I think." That's a dangerous fallacy. And it's a fallacy Congress itself tends to perpetuate, because in today's polarized environment, politicians on all sides like to claim that their views are the only ones backed by science.

That's not to say that Congress doesn't ever simply ignore science. But the idea that Congress would have acted more responsibly about, say, climate change if the OTA had been around is absurd. Congress has not lacked information about climate change. Scientists can lead congressmen to information, but can't make them think. Otherwise, the 2001 National Academy of Sciences climate report completed at the request of President George W. Bush would have been enough to alter the climate debate in Congress.

But climate change is an atypical issue in that for so long Congress was debating a purely scientific question — is climate change real? — and was largely impervious to an unusually broad scientific consensus. Now that Congress is asking for scientific advice on climate, it may be about to err in the opposite direction.

A House spending bill would hand over to scientists the decision on how to allocate \$45 million among federal agencies for climate adaptation and mitigation. Specifically, the decisions would be made by a panel chaired by the president of the National Academy of Sciences — who would also choose six of its members — and would include the heads of seven federal agencies and the president of the National Academy of Engineering. The Senate counterpart to the bill does not have a similar provision, and it's too early to know whether Congress will ultimately create the panel.

But the proposal sets a bad precedent. How to spend money on climate adaptation is not a purely scientific question, and the federal government should not be delegating its decision-making on it to a group of private scientists (who would be a majority on the commission) with no accountability to the public. Moreover, leadership of the panel by the National Academy of Sciences would make the academy seem biased if it later evaluated the resulting programmes, which is its traditional role. Worse still, the academy would be likely to tarnish its reputation as an expert arbiter of scientific disputes related to policy decisions if it started to make policy decisions itself.

Congress should do everything possible to hear from scientists. That's why re-establishing the OTA is a worthy idea even though its consequences would be less far-reaching than many scientists think. But Congress and scientists need to stop pretending that science alone can determine policy decisions. Claiming that a particular policy approach is the only one justified by science simply opens science to attack, while muddling discussion of the values issues at stake. In the long run, Congress may do more harm to science by setting up the proposed climate adaptation panel than it ever did by closing down the OTA. ■

David Goldston is a visiting lecturer at Princeton University's Woodrow Wilson School of Public and International Affairs.



Blended wing-body technology, as tested in Boeing's X-48B, is one option for making aircraft more efficient.

R. FERGUSON/BOEING

Civil aviation faces green challenge

As a fast-growing emitter of greenhouse gases, the aviation industry is under intense pressure to improve its fuel efficiency. **Kurt Kleiner** surveys its options.

Last Sunday, Boeing rolled out its latest airliner, the 787. It boasts 20% improvements in fuel efficiency, and is the first large airliner whose main structure and skin is built of light, carbon-fibre composites, instead of metal alloys. But the industry, under pressure over its role in global warming, is scrambling to improve efficiency yet further.

"Our goal is to take the industry to the next level, with respect to environmental responsibility," says Anthony Concl, a spokesman for the International Air Transport Association.

At last month's Paris Air Show, this effort took centre stage. Airbus called for global collaboration on the technical challenges ahead. Giovanni Bisignani, director-general of the International Air Transport Association, said there should be a "zero-emissions" plane by 2050.

Civil aviation accounts for about 3.5% of total greenhouse-gas emissions, according to the Intergovernmental Panel on Climate Change, but this could rise as air travel becomes more popular. And the pursuit of greater fuel efficiency is also being driven by airlines' fuel bills, given the high price of oil. "We know that in order to be able to market a new aircraft, it must be between 15% and 25% better in fuel consumption than the aircraft it's going to replace," says Rainer von Wrede, director of environmental affairs at Airbus.

One place that Boeing and Airbus, which between them build all of the world's large civilian airliners, look for such improvement is in aeroengines. Today's civilian jets use turbofans, in which a gas turbine at the rear of

the engine drives a fan mounted just in front, which provides most of the thrust. A turbofan is quieter and more energy efficient than a pure gas-turbine engine, in which the thrust comes entirely from the jet of hot air shooting out the back. Efficiency will be increased if these fans are made bigger. But as they expand, so do the ducts that surround them, and their increased weight and drag soon cancels out the advantage of the bigger fan.

In an EcoJet proposal also unveiled at Paris, British budget airline EasyJet suggested a plane that would halve carbon dioxide emissions per passenger mile by 2015. It proposes using open rotors, in which the duct is eliminated. The bigger fans that this makes possible could create an engine that is 15% more efficient, says David Clarke, head of technology development at engine-maker Rolls-Royce in Derby, UK. But eliminating the duct increases noise — and leaves less protection for the cabin in the event that a blade fails. EasyJet suggests positioning the turbofans above and behind the plane, where the tail and body would block the noise to the ground, and where a broken blade would miss the cabin.

A similar increase in efficiency could come from a geared turbofan. In a conventional turbofan, the gas turbine in the back of the engine turns the fan directly by means of a shaft. The fan is most efficient, though, at a much slower rotational speed than the most efficient speed for the turbine. Conventional engines are

forced to operate at a compromise speed.

Pratt & Whitney, a jet-engine manufacturer based in Hartford, Connecticut, is already testing a new engine that places a gearbox between the turbine and the fan. It allows the fan to run at a third of the speed of the turbine, and will give a 15% increase in efficiency, says Bob Saia, head of next-generation products there.

Beyond the engines, airliner makers are looking to better body designs to improve fuel efficiency. The simplest improvement is to make existing body designs lighter. The

Boeing 787, which is expected to fly commercially next year, is leading the way in this regard, using carbon-fibre composites for its entire airframe. Stronger composites, reinforced by

carbon nanotubes, are widely regarded as the next step.

Engineers are also using computer simulation to further reduce drag from airliners. The wings on the latest version of the Boeing 737 generate 30% less drag than the ones they replaced, says Bill Glover, director of environmental performance strategy for Boeing in Seattle, Washington. Increasingly powerful simulation tools promise even better designs, adds Hans Weber, an aviation analyst with Tecop International in San Diego, California. This allows engineers to vet sophisticated new designs thoroughly on computer, before embarking on the much more costly business of building prototypes for tests in wind tunnels.

Additionally, manufacturers are toying with

"Increasingly powerful simulation tools promise even better designs."

radical body concepts, such as the blended wing-body that was pioneered in the US Air Forces' B2 stealth bomber. In these designs the tube-like fuselage and tail are eliminated, turning the entire aircraft into a lift surface. A blended wing-body design currently being ground-tested by Boeing and NASA (pictured) promises a 20% fuel reduction. But industry watchers are unsure if passengers would accept such a design: they'd have to sit in a wide, windowless compartment, and those near the wing ends would be shaken up and down whenever the airliner banked. Airport gates would also have to be entirely redesigned to accommodate such a plane.

Fuelling change

The industry is also looking at how new fuels could help it to assuage environmental concerns. Biofuels have obvious attractions — provided the public accepts that their growth and combustion are generating no net carbon emissions. Bioethanol, the most popular bio-fuel for car engines, will freeze at high altitudes, but biodiesel and biobutanol could each be burned in conventional turbofan engines. Richard Branson, chairman of London-based Virgin Airlines, pledged to fly a commercial jet on biofuels in 2008, and is working with Boeing and the aeroengine division of General Electric to test candidate fuels.

The ultimate zero-emissions fuel may prove to be hydrogen, which could easily be burned by conventional jet engines with no carbon emissions. First, the world will have to figure out how to make and distribute hydrogen cheaply and safely. And even given that, planes would require a redesign. Although liquid hydrogen provides three times the energy of the same weight of kerosene, it takes up seven times as much space. That would require aircraft with bigger fuel tanks, as well as the insulation to keep the liquid cool.

Once a plane exists with the cooling equipment that allows it to use liquid hydrogen as fuel, the potential exists for a more efficient all-electric plane, says Gerald Brown, a physicist at NASA Glenn Research Center in Cleveland, Ohio. Conventional generators are too heavy to be efficient on aircraft. But with a source of cold hydrogen, a superconducting generator and electric motors start to make sense, he argues, and might be more efficient than today's turbofans (P. J. Masson *et al.* *Supercond. Sci. Technol.* **20**, 748–756; 2007).

In the end, the aviation industry will seek to combine all of these approaches in a concerted bid to head off accusations that its rapid growth is spurring global warming. "We have a lot of creative people with a lot of great ideas. I'd say nothing's off the table," says Boeing's Glover. ■

IN BRIEF

BRAZIL DEAL Abbott Laboratories, the Illinois-based drug company, reached agreement with the Brazilian government to sell its AIDS drug Kaletra there for a reduced price — lifting the threat that Brazil might license it compulsorily. On 4 July, the company said it would sell it there for \$1,000 per patient per year, down by about 30%. Abbott has drawn criticism from AIDS activists for its approach to marketing Kaletra in Thailand, where it has withdrawn an application to sell a heat-stable version of the drug (see *Nature* 448, 14; 2007).

ALL IN HAND A 2,000-page guidebook has been compiled to help specialists in public health and agriculture to navigate their way through the thickets of intellectual property. The *IPHandbook* (www.iphandbook.org) has been produced by two non-profit organizations that specialize in opening up access to intellectual property in health and agriculture, respectively, as a resource for scientists and government officials, especially in poor countries. Support from the Rockefeller Foundation and other sponsors will permit its free distribution in low-income nations.

FLOAT ABANDONED A company seeking to develop medical technologies at the University of Liverpool, UK, withdrew plans for an initial public share offering on London's Alternative Investment Market (AIM). ULive, which has been seeking to exploit intellectual property developed at the university's labs, had hoped to raise £20 million (US\$40 million) in late June, in a flotation that would have valued it at £70 million. But the number of initial public offerings on the AIM is down by half on last year, and lack of investor interest in recent offerings from small companies led ULive to withdraw its plan.

MARKET WATCH



This week Wood Mackenzie, an Edinburgh-based research and consulting firm, reviews recent trends in biotechnology stocks.

The Nasdaq biotechnology index experienced a choppy ride — overlying a general, downward trend — during May and June. The index lost more than 4% of its value, although it is still up nearly 2% on the start of 2007. By contrast, broader market indices made modest gains during the same period.

As it is weighted according to market capitalization, the biotechnology index is closely tied to the fortunes of its largest component company, Amgen. The Californian company has suffered from concerns in the United States over the safety of erythropoiesis-stimulating agents in cancer treatment, of which Amgen has two: Epogen and Aranesp. Investors reacted badly to the threat to Amgen's largest franchise, which accounts for about half of the company's sales. The company's share price fell,

and was down nearly 12% over the period (see *Nature* **447**, 899; 2007).

Other companies in the cancer market have had misfortunes, too. In early May, the US Food and Drug Administration (FDA) said it would require more data before approving a therapeutic prostate cancer vaccine, Provenge, under development by Seattle-based Dendreon. Its shares fell more than 60% on the news, wiping out gains made in March. And in early June, the FDA temporarily halted all clinical trials for the ovarian cancer drug candidate Telcyta, on safety grounds. Shares of its developer, Telik of Palo Alto, California, lost almost half their value on the news.

Summer is customarily a slow time for stock markets in general, and biotechnology shares in particular. This year, Amgen's misfortunes are unsettling investors, with many of the smaller companies in the biotech sector involuntarily joining in the slide. ■

Reading the book of death

Studies of mass extinctions tend to emphasize the sheer scope of the carnage. But subtle differences between the species that died and those that survived can be crucial, finds **Nick Lane**.

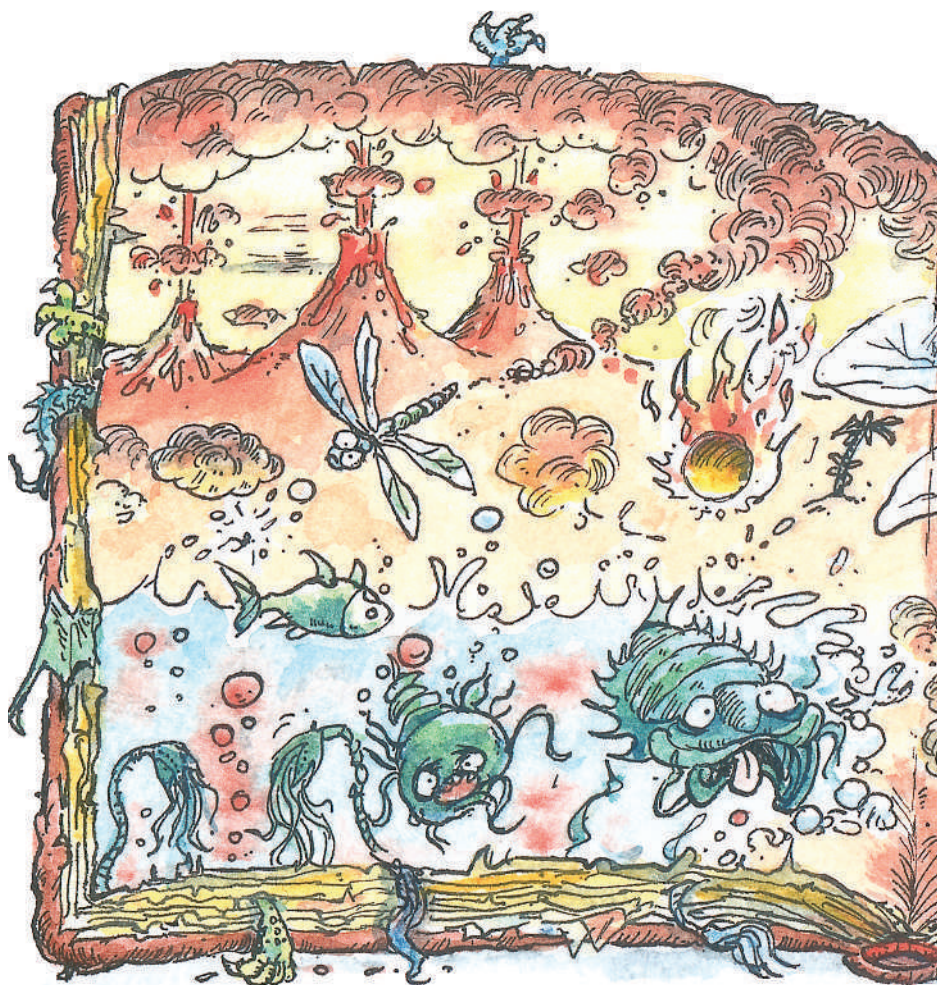
The extinction at the end of the Permian period, some 251 million years ago, is the most fascinating mass-murder mystery in Earth's history. Simply put, a party or parties unknown killed off up to 96% of all species then alive. Palaeontologists and Earth-system scientists have suggested any number of possible culprits, from a comet or asteroid collision to a collapse of the ozone layer, from a dramatic suite of volcanic eruptions to ocean waters bubbling with sulphurous fumes.

Given that the evidence for an impact, although championed by some, seems sketchy and inconclusive to the majority, the most dramatic of these suspects is volcanism — the sequence of eruptions that gave rise to the basalts of the Siberian Traps. The scale of the eruptions was vast, with something like 3 million cubic kilometres of basalt flowing on to the surface, and the main lava flows took place at almost exactly at the same time as the end-Permian extinction itself, give or take a few hundred thousand years. The vast clouds of gas and ash they spewed out look like the smoke from a very large gun indeed¹.

But if the erupting traps look increasingly like the trigger for the great die-off, how did they actually do their damage? The fiery birth of a small continent's worth of new landscape can change everything from the brightness of the sky to the chemistry of the oceans. How can scientists work out which of these effects were inconveniences, and which were fatal?

One answer is to study not the volcanoes, nor their victims, but the creatures that survived. A mortality rate of 96% sounds pretty indiscriminate — but recent physiological comparisons between the creatures that died and those that survived reveal intriguing patterns². And these patterns do not just provide clues as to how the great dying actually unfurled. They also reveal how the extinctions at the end of the Permian period shaped the subsequent history of the world that the survivors inherited.

Until recently, perhaps the most popular 'kill mechanism' evoked for the extinction was global warming³. Volcanic carbon dioxide, so the story goes, raised temperatures enough to destabilize vast reserves of methane hydrates in the sea floor, pumping temperatures up further. But although there's little question that global temperatures rose at the end of the Permian by as much as 6 °C, there are problems with this picture.



Among the things to go extinct at the end of the Permian are a lot of plants. There are plausible arguments as to how temperature could kill off animals — for example, it would force a higher metabolic rate on them, requiring an increase in their calorie intake that they might simply not have been able to satisfy. But there's no obvious reason why such heat would damage plants. What's more, the physiology of the survivors doesn't fit the story. Animals that seem to have started off with higher metabolic rates suffered less than those with slow metabolisms. Other things being equal, you might expect animals that are already generating more heat in their own bodies to do worse when things heat up further.

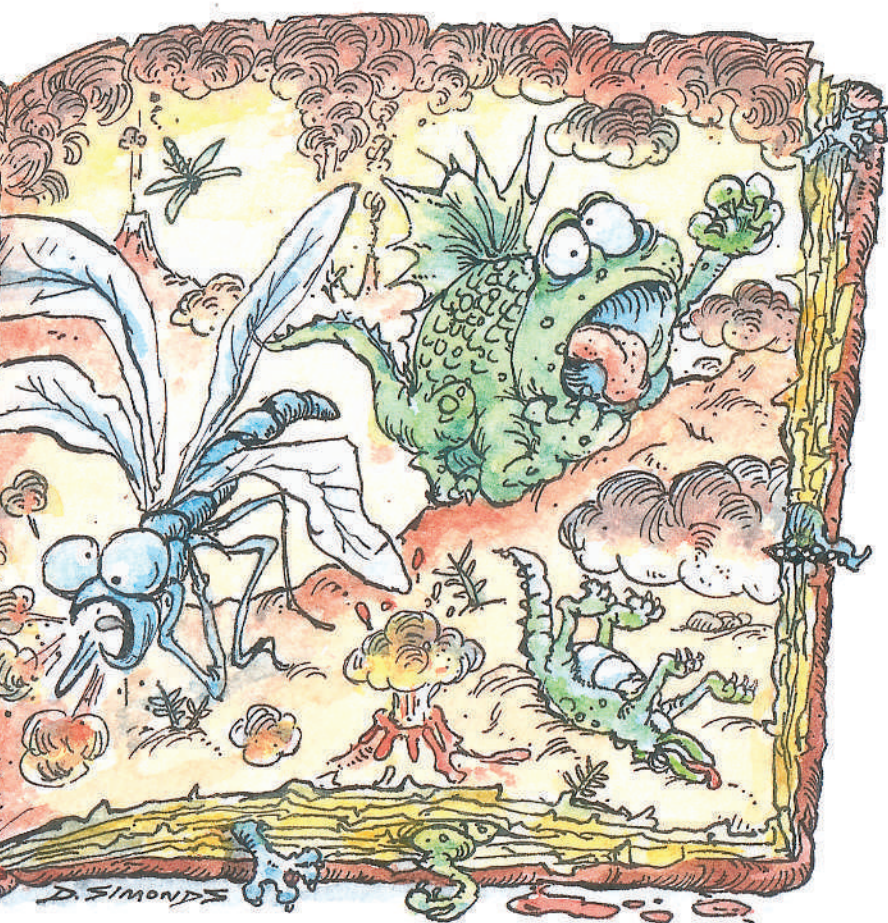
Another possible kill mechanism is a collapsed ozone layer. Mutant pollen and spores in sediments from the time of the extinction hint at an increase in ultraviolet levels brought about by a dearth of ozone. David Beerling, a plant physiologist who concentrates on palaeoclimate issues at the University of Sheffield, UK, says his eyes were opened to this risk when he learned that a NASA aircraft flying through the plume above the 2000 Hekla eruption in Iceland showed substantial amounts of hydrochloric acid had been injected directly into the stratosphere. Simulations by Beerling and his colleagues at Sheffield and the University of Cambridge, UK, suggest that the Siberian Traps could have removed 70% of the ozone shield from the Northern Hemisphere, enough to do serious environmental damage⁴.

Even so, Beerling doubts that an ozone collapse was responsible for the mass extinction: "Animals have too many places to hide on land, and in the oceans they could

"Animals have too many places to hide on land, and in the oceans they could easily escape ozone collapse."

— David Beerling

D. SIMONDS



easily escape the worst of it," he says. And as with global warming, an excess of ultraviolet radiation does not seem to fit the pattern of survival that is actually seen. Living deep in the sea doesn't seem to have helped.

But better ways of breathing, which are often seen in creatures with a higher metabolic rate, did. In the sea, creatures with gills and an active internal circulation, such as snails, clams, crabs and fish, fared better than stationary filter feeders. A similar pattern holds on land, with groups that were not so good at pumping air through their respiratory systems — such as the giant dragonflies iconic of pre-Permian times — suffering disproportionately. These patterns focus attention not on harsher sunlight, but on tainted air.

Oxygen collapse

One possible factor is a worldwide dearth of oxygen. It used to be thought that today's oxygen level of 21% had remained fairly constant over geological time. But in the late 1980s, Robert Berner of Yale University in New Haven, Connecticut, shocked the establishment by arguing that in the Carboniferous period that preceded the Permian, 300 million years ago, the oxygen level had reached 30% or more, and that it had then fallen dramatically to a mere 13% in the late Permian and the early part of the subsequent Triassic. But after a few refinements, Berner's models for this period have been widely accepted.

Berner blames the oxygen collapse on another spate of volcanism that predates the Siberian Traps by eight million years or so. The warming those eruptions triggered, he argues,

would have brought about more arid conditions, drying out some of the coal swamps of the time. It had been the capacity of those swamps to bury organic carbon where it could not be respired back into carbon dioxide, which had accounted for the high oxygen levels. As the swamps dried out, less organic carbon got buried this way, and so more was respired back to carbon dioxide — using up oxygen in the process. Over millions of years the slower rate of carbon burial drew oxygen levels down to an unprecedented low point⁵.

The falling oxygen levels Berner describes would have restricted the geographical range of many land animals⁶. Most animals have an 'altitude ceiling' above which they can't flourish (for humans it's 5.1 kilometres). As oxygen levels dropped in the later Permian, so these ceilings would have pressed down on the creatures beneath them. According to palaeontologist Peter Ward of the University of Washington in Seattle, for creatures that had evolved in a high oxygen world, modest hills of just a few hundred metres could have become impassable barriers.

In the oceans, too, the habitable zone would have shrunk as oxygen levels fell. Global warming would have made matters worse, as oxygen is less soluble in warm water than in cold. Such effects are still visible in the world today. This January a study of fish populations in the North Sea showed that, among eelpouts (Zoarcidae), the size of both populations and individuals shrinks when increased temperatures lessen the oxygen supply⁷.

Palaeontologist Richard Twitchett, at the University of Plymouth, UK, thinks that something similar happened in Permian times. His evidence comes from the species that seem to vanish at the end of the Permian, only to start cropping up again after a hiatus of millions of years — the 'Lazarus taxa'. A startling 70% of the gastropods found in the mid-Triassic seem to have 'risen from the dead' in this way. Twitchett thinks that their populations became so small that random chance expunged them from the fossil record. Things didn't pick up for the gastropods until oxygen levels started to pick up, well into the Triassic⁸. And when they did re-emerge, they were smaller — they have something of Lilliput, as well as Lazarus, about them.

But for all this it is hard to pin the mass extinction proper on oxygen. From a physiological point of view, the species most likely to survive low oxygen would seem to be those



Extinction victim: *Dinogorgon rubidgei*, a mammal-like reptile.

J. BLAIR/CORBIS

with a low metabolic rate — and yet the stationary filter feeders, which fit that bill, fared disastrously. And then there's the timing. The trends in atmospheric oxygen levels were slow, reflecting changes in the rate at which organic matter was buried over millions of years. By contrast, detailed analyses of the strata in which the fossils are preserved show that the extinction was abrupt, with most species disappearing in a few tens of thousands of years⁹. As Andrew Knoll, a palaeontologist at Harvard University in Boston, Massachusetts, puts it: "If you've got gradual changes taking place over millions of years, I'd expect organisms to adapt; but if the changes take place in 10,000 years, I'd bet on the environment to club them."

Killer carbon

Low oxygen must have placed life in a precarious position, but something else pushed it over the edge. And a good clue to the culprit is a series of unusual shifts in carbon isotopes found in strata worldwide, corresponding closely to the exact time of the extinction.

This series of spikes in the carbon-isotope level suggests that strange things were happening to the carbon cycle at this time. The anomalies — sudden upswings in the ratio of lighter carbon-12 to heavier carbon-13 — are too big to be explained by changes in the total amount of biomass (which tends to be carbon-12 rich). The most plausible explanation would be bursts of methane. But sea-floor methane hydrate reserves would not be up to the job. There are five distinct spikes in the record between the latter part of the Permian and the early Triassic. According to Jonathan Payne at Harvard University and his colleagues, although sea-floor hydrate reserves might account for one such spike, they could not be replenished quickly enough for a whole succession¹⁰.

Greg Retallack of the University of Oregon in Eugene points to a different source. In both the Siberian Traps and the earlier volcanism, lava rose to the surface through both carbonate rocks and coalfields. Hot lava passing through coal measures breaks down some of the heavy hydrocarbons into methane, and although this 'thermogenic' methane is not as light, isotopically speaking, as the stuff on the sea floors, it is light enough to explain the isotope spikes¹¹.

At a time when oxygen levels were already stressing animal life, such large injections of methane, which would oxidize into carbon monoxide and carbon dioxide, could have killed animals through simple asphyxiation. As for the plants, Retallack points out that those that did the worst were in swamps, which were low in oxygen anyway, so their roots asphyxiated.

The need to deal with high levels of carbon dioxide could explain various aspects of the survivors' physiology, including subtleties at the molecular level. Knoll is intrigued by a subtle difference in pigment use between the squid-like ammonoids, which died out, and the creatures from which today's spiral-shelled nautilus is descended. Animals use

"Even if death doesn't get you right away, you're unlikely to have much spare energy for sex."

respiratory pigments such as haemoglobin to get oxygen from the lungs or gills to the tissues that need it, and to remove the resulting carbon dioxide. When carbon dioxide levels get too high, some such pigments will cease to work properly, unable to remove the bodily carbon dioxide with which they are encumbered or to pick up more oxygen. The nautilus, Knoll notes, has a pigment that deals well with high levels of carbon dioxide. The closest living relatives of the ammonoids, the coleoids — soft-bodied cephalopods such as squid and octopus — do not.

Their unhelpful pigments might not have been enough to doom the ammonoids — but as shell makers they had other problems too. One of the effects of increasing carbon dioxide is that sea water, like soil, becomes more acid. Shell formation, Knoll argues, is repressed if carbon dioxide levels cause organisms to lose control over their systems for dealing with such acidity. The groups least good at regulating acidity seem to have been particularly badly hit at the end of the Permian. So were those creatures that needed their shells to channel water over their gills.

Sulphur survivors

The case for a significant amount of asphyxia is strong. But global warming and low oxygen would have brought another deadly gas into play. Hydrogen sulphide is produced in vast quantities by bacteria that gain their energy from sulphate, and that thrive only in very low oxygen conditions. Today, these conditions are prevalent only in closed basins, most strikingly that of the Black Sea. But there's plenty of evidence to suggest that such conditions applied much more widely at the end of the Permian.

Lee Kump, a geochemist at Pennsylvania State University in University Park, argues that sulphidic deep waters could have crept higher and higher up the water column, killing everything they encountered and creating what is evocatively named a 'Strangelove ocean'. When they reached the surface they would have released poisonous gas straight into the air¹². Last year, biomarkers unique to 'green sulphur' bacteria, which make use of sulphide for photosynthesis and cannot tolerate oxygen, were found in some sediments from the end of the Permian, suggesting that oxygen-free waters got close enough to the surface for photosynthesis to take place¹³.

The effects of hydrogen sulphide escaping from the oceans would in all likelihood have been local, not global, and Beerling's models suggest that no plausible level of the gas could have damaged the ozone layer, as had been suggested. But it would have been another challenge for stressed ecosystems.

So with dangerously low atmospheric oxygen levels, ecosystems were compressed and fragmented. The deep oceans were largely uninhabitable. Land plants were dying back in the arid greenhouse climate, making food hard to come by. And then came the hammer blows of fate, the greatest volcanic outpourings in the history of our planet, releasing vast quantities of methane and carbon dioxide, raising global temperatures by 6 °C, and lifting the



Eruptions, such as Hekla's, could cause ozone depletion.

Strangelove conditions to the very surface.

Head for the hills and there's no oxygen; stay on the shores and you risk breathing hydrogen sulphide. High carbon dioxide levels sabotage your respiratory pigments and choke you from within. Even if death doesn't get you right away, you're unlikely to have much spare energy for sex. Population sizes fall; so do body sizes. Even for those that survive the immediate toxicity, slow extinction was likely over a few generations — the blink of an eye in geological terms.

And the resultant catastrophe shaped the rest of the planet's history. According to research by Peter Wagner and his colleagues at the Chicago Museum of Natural History in Illinois, life in the seas before the extinction was split roughly 50–50 between simple marine ecosystems dominated by sessile filter-feeders, such as today's sea lillies, and more complex systems in which larger numbers of species moved around and interacted. After the extinction and the subsequent recovery, complex ecosystems outnumbered simple ones by three-to-one¹⁴. That has remained pretty much the case ever since, the patterns of biodiversity reflecting the extinction bottleneck that a few creatures squeezed through with frantically fluttering gills.

The situation was similar on the continents. The hardest group of reptiles was the cynodonts, from which mammals are descended. The squat, flat-faced *Lystrosaurus*, with its breathing aided by a bony palate separating the mouth from the nose and a barrel-chest with a muscular diaphragm to help deep breathing, went on to dominate the Triassic. Cynodonts also enjoyed the benefits of nasal turbinates.

Some turbinates are linked with fast metabolic rates in warm-blooded creatures, as they restrict water loss during rapid respiratory ventilation. In the early Triassic, though, they might have restricted water loss in hyperventilating animals on the verge of suffocation¹⁵. Other survivors included the archosaurs, which benefitted from breathing made easier by air sacs. Their descendants became the dinosaurs, birds and crocodiles¹⁶.

It's unlikely that any of these respiratory adaptations evolved in response to the suffocating conditions at end of the Permian; rather, the animals most likely to survive the disaster were those already adapted to deal with suffocating conditions on a daily basis — creatures used to burrowing on land, or scavenging in the stagnant muds of ocean shelves; or just those that came up trumps with an air-sac system that happened to be particularly good.

Take a deep breath, fill your lungs with a toxic brew of gases, gasp for the thin oxygen. Small wonder most life perished; wouldn't you? But your ancestors survived and repopulated the world in new ways that we still see



"Head for the hills and there's no oxygen; stay on the shores and you risk breathing hydrogen sulphide."

around us today. The genetic memory of those times is etched in our physiology — and in the living fabric of our world.

Nick Lane is the author of *Power, sex, suicide: Mitochondria and the meaning of life*.

1. Kamo, S. L. *et al. Earth Planet. Sci. Lett.* **214**, 75–91 (2003).
2. Knoll, A. H., Bambach, R. K., Payne, J. L., Pruss, S. & Fischer, W. W. *Earth. Planet. Sci. Lett.* **256**, 295–313 (2007).
3. Benton, M. J. & Twitchett, R. *Trends Ecol. Evol.* **18**, 358–365 (2003).
4. Beerling, D. J., Harfoot, M., Lomax, B. & Pyle, J. A. *Phil. Trans. R. Soc. B* **365**, 1843–1866 (2007).
5. Berner, R. A. *Geochim. Cosmochim. Acta* **69**, 3211–3217 (2005).
6. Huey, R. B. & Ward, P. D. *Science* **308**, 398–401 (2005).
7. Pörtner, H. O. & Knust, R. *Science* **305**, 95–97 (2007).
8. Twitchett, R. J., Krystyn, L., Baud, A., Wheeley, J. R. & Richoz, S. *Geology* **32**, 805–808 (2004).
9. Jin, Y. G. *et al. Science* **289**, 432–436 (2000).
10. Payne, J. L. *et al. Science* **305**, 506–509 (2004).
11. Retallack, G. J. & Krull, E. S. in *Wetlands Through Time* (eds Greb, S. & DiMichele, W. A.) 249–268 (Geological Society of America, Boulder, Colorado, 2006).
12. Kump, L. R., Pavlov, A. & Arthur, M. A. *Geology* **33**, 397–400 (2005).
13. Grice, K. *et al. Science* **307**, 706–709 (2006).
14. Wagner, P. J., Kosnik, M. A. & Lidgard, S. *Science* **314**, 1289–1292 (2006).
15. Berner, R. A., VandenBrooks, J. M. & Ward, P. D. *Science* **316**, 557–558 (2007).
16. O'Connor, P. M. & Claessens, L. P. A. M. *Nature* **436**, 253 (2005).

D. SIMONDS



A CULTURE IN THE BALANCE

Traditional Chinese medicine and Western science face almost irreconcilable differences. Can systems biology bring them together? **Jane Qiu** reports.

Liu Wen-long's modest Beijing practice looks no different from most clinics. But he is no ordinary doctor. Liu never orders lab tests, nor does he prescribe high-tech imaging diagnostics. He relies on simple observations, checking a patient's pulse, complexion and odour, and asking about habits and medical history. At 69 years old, he has been practising traditional Chinese medicine for 43 years and he is resolute about its benefits. "People keep coming back because it cures them and improves their well-being," he says.

Indeed, patients trickle in to see Liu all morning for conditions ranging from allergies to lung cancer. Some are nervous first-timers, others are regulars, confident in what traditional Chinese medicine has to offer. Ms Huang, an accountant from the outskirts of Beijing, is delighted that her migraines, which haunted her for years, disappeared after three herbal regimens. "I used to live on painkillers and felt tired all the time," she says. "I am now a totally different person."

In a country that is fiercely embracing modernity, clinics such as Liu's, which have been operating the same way for thousands of years, seem vulnerable and out of place. Indeed, attitudes on traditional Chinese medicine have divided the country. Last year, Zhang Gong-yao, from the Central South University in Changsha, Hunan, published an article in a Chinese journal calling traditional Chinese medicine a pseudoscience that should be banished from public healthcare and research¹. The article caused uproar in the country, and earlier this year the government announced an ambitious plan to modernize the millennia-old practice².

But should such a formidable gap be bridged? Modern Western medicine generally prescribes treatments for specific diseases, often on the basis of their physiological cause. Traditional Chinese medicine, however, focuses on symptoms, and uses plant and animal products, minerals, acupuncture and moxibustion — the burning of the mugwort herb (*Artemisia*

vulgaris) on or near the skin. But whether these methods are effective and, if they are, how they work remain a source of some derision. The greatest divide is in the testing. In the West, researchers test a drug's safety and efficacy in randomized, controlled trials. Traditional Chinese treatments are mixtures of ingredients, concocted on the spot on the basis of a patient's symptoms and characteristics and using theories passed down through generations.

The mainstream medical community, in China and abroad, has been highly critical of the underlying theories. Traditional Chinese medicine is based on ideas such as *qi* (meridian), in which illness is caused by blocked energy channels; *yin* and *yang*, which emphasizes the balance of energy; and *wuxing* (five elements), in which people's organs and health status are categorized according to their 'elemental characteristics': fire, wood, water, earth and metal.

Pharmaceutical companies have become more interested in traditional Chinese

P. HILTON/EPA/CORBIS

medicines over the past decade. But their approach has been characteristically Western: isolate the active ingredients and test them one at a time. This reductionist approach has led to the approval of drugs such as artemisinin for malaria, which is used to treat fever in traditional Chinese medicine, and arsenic trioxide, which has been carried over from Chinese medicine for treatment of acute promyelocytic leukaemia.

But identifying the active ingredients isn't easy. Most remedies in traditional Chinese medicine, as it turns out, are compound formulae — or *fufang* — that contain as many as 50 species of herbs, and thousands of chemicals therein (see 'Knowledge mining'). To tap into the deeper well of traditional Chinese treatments, researchers think they may need to look at how the mixtures of ingredients act in concert.

Relaxed regulation

The criteria for approval of herbal mixtures as medicines are now starting to relax, at least in the United States. In June 2004, the US Food and Drug Administration (FDA) issued new guidelines that permit the approval of herbal mixtures if they can be shown to be safe and effective, even if the active constituents are not known. Last October, the FDA approved the first such botanical drug under the new rules, a proprietary mixture of green-tea extracts called Veregen developed by the German company MediGene for treating genital warts.

These new regulations have helped to renew industry's interest in the complex formulae. And a buzzing new Western field could be poised to capitalize on the deeper secrets of traditional Chinese medicine. Systems biology attempts to understand the function and behaviour of an organism by studying the interactions between its components. It has been called a more holistic approach to biology and is seen by some as a perfect match for traditional Chinese medicine.

By measuring many genes, proteins and metabolites at the same time, systems biology may provide a measure of the entire body's response to a complex mixture of herbs. "If there is any technology that could lead to a breakthrough in traditional Chinese medicine, it will be systems biology," says Robert Verpoorte, head of the pharmacognosy department at the University of Leiden in the Netherlands. But not everyone agrees that the new technology is equipped to test old ideas.

Jia Wei, a pharmacologist at the Shanghai Centre for Systems Biomedicine at Jiao Tong University, and Tang Hui-ru at the Wuhan Institute of Physics and Mathematics, part

Knowledge mining

The key to Chinese herbal medicine is to choose a combination of plant species based on the particular symptoms and characteristics of the patient and guided by the various theories of traditional Chinese medicine. There are many schools of thought on the causes of disease and the best way to formulate herbal medicines to tackle them.

One of the most influential theories is the principle of *jun-chen-zuo-shi*. The *jun* (emperor) herbs treat the main cause or primary symptoms of a disease. The *chen* (minister) herbs serve to augment or broaden the effects of *jun*, and relieve secondary symptoms. The *zuo* (assistant) herbs are used to modulate the effects of *jun* and *chen*, and to counteract the toxic or side effects of these herbs. The *shi* (courier) herbs are included in many formulae to ensure that all components in the prescription are well absorbed, and to help deliver or guide them to the target organs.

For researchers



SALVIA MILTIORRHIZA

aka: *Danshen*

With their blood-thinning qualities, remedies containing roots from this plant are traditionally used to treat cardiac and vascular disorders. Some animal studies suggest uses in liver and kidney disease.



PANAX GINSENG

aka: Asian ginseng

Used to improve mental acuity, and to treat senile dementia. Often combined with ginkgo biloba. Some evidence suggests blood-sugar lowering effects for type 2 diabetes.

trying to develop clinically tested drugs from Chinese herbal medicine, one of the most challenging steps is scouring the extensive ancient and contemporary literature to produce a short list of compound formulae that might be promising for the disease under study. More than 11,000 plant species used in more than 100,000 compound formulae have been recorded in China, and the same plant species can have distinct roles in different formulae. Thus, *jun* and *chen* herbs in one formula may serve as *zuo* and *shi* herbs in another.

It can also be difficult to translate ancient interpretations of disease into those used in modern medicine. For example, many herbs are said to have the function of *qingre* (to clear heat). What this really means is that they may have antibacterial, antiviral or anti-inflammatory effects. Similarly, a guiding principle of traditional medicine practice is *fuzheng quxie* (to help the righteous and combat the evil), which, in the language of Western medicine, is to boost the immune system and evade viruses. "It's all about local knowledge," says Robert Miller, chief executive of Phynova in Oxford, UK, a pharmaceutical company that develops Western-styke drugs from medicines identified from China. "Having in-depth understanding of traditional Chinese medicine is absolutely essential." **J.Q.**

of the Chinese Academy of Sciences, want to understand more fully how herbal extracts affect the whole body. They are collaborating with Jeremy Nicholson, head of the department of biomolecular medicine at Imperial College London, and using technologies such as nuclear magnetic resonance spectroscopy and mass spectrometry to profile the metabolites in a person's urine or blood — a discipline they call metabonomics.

Jia and his colleagues found that rats given the compound 1,2-dimethylhydrazine to induce tumours in their colons had different metabolic profiles in their urine from those in the control group. And by feeding the rats a combination of two herbal extracts — *Coptidis rhizoma* and *Evodia rutaecarpa*, which are widely used in traditional Chinese medicine to treat gastric conditions — the researchers were able to reverse these changes in metabolism.



Mix and match: traditional Chinese medicine uses a variety of products to treat people's symptoms.

Their results have not yet been published, but the researchers say that by looking at the changes in metabolites in detail, they have pinpointed the metabolic pathways that the herbs affect.

Culture shift

Jan van der Greef from SU Biomedicine in Zeist, the Netherlands, and his colleague Wang Mei are using a similar approach. In a mouse model of metabolic syndrome — a cluster of conditions such as insulin resistance and high blood pressure that often occur together — they and their team looked at the effect of an undisclosed formula used in traditional Chinese medicine on lipid profiles. When these mice are fed a high-fat diet, they become more resistant to insulin. The lipid profiles of these mice were clearly distinguishable from those of mice fed a normal diet, and they shifted towards the healthy state when the mice were given traditional Chinese medicine³.

The researchers noticed that the profile shift resembled that caused by the Western obesity drug Rimonabant, which acts on proteins called CB-1 endocannabinoid receptors. And their unpublished work with cell culture suggests that herbal extracts can affect lipid metabolism through the same receptor, says van der Greef. The team is now testing the formula in clinical trials.

Although one active ingredient may act as the Western drug, the uncertain role of additional ingredients and the variability of active ingredients confounds Western sensibilities. "Variations worry people," Nicholson says. The same plant species grown in different regions and harvested in different seasons could have distinct chemical compositions. This has always been a vexing issue for herbal-medicine researchers.

While at Nicholson's lab, Tang and his colleagues analysed the molecular components in chamomile plants from Egypt, Slovakia,

and Hungary, and could classify them easily⁴. Using similar approaches, the team from the Wuhan Institute of Physics and Mathematics found significant variations in the same herbal medicines produced by different companies and even between different batches produced by the same company. "This is an issue China must tackle for its herbal medicines to raise their game in the world market," says Tang.

To many self-purported systems biologists, several approaches are needed to build a complete picture of a living organism and to understand the effect of traditional Chinese medicine. Nevertheless, systems biology has been a conspicuously hard field to define. Many have used the term loosely, and pioneers in the field contend that the technologies haven't been honed to the point that they could be used for these approaches.

"It's conceivable that systems biology could find applications in trying to sort out components in Chinese herbal medicine, but it's very early days," says Leroy Hood, president of the Institute for Systems Biology in Seattle, Washington, and regarded as the field's founding father. "It would be an enormous challenge at this point and time."

Systems biology has been successful in model organisms, according to Hood, but is much less so in human studies. Many hurdles need to be overcome before researchers could even begin to contemplate how to deal with subjects as complex as traditional Chinese medicine. For example, better detection systems are needed to measure metabolites, especially proteins, accurately in the blood, and more powerful computational and statistical tools are crucial for dealing with large and complex data sets. "Those technologies are at early stages of maturation," Hood says.

There are also broader concerns about the modernization of traditional Chinese medicine, from both advocates and sceptics of the practice. Some are uncomfortable with separating the study and development of Chinese herbal medicines from the theories that underlie its normal practice. "Traditional Chinese medicine is not just a medical system, but a branch of philosophy and healing art that is an important part of Chinese culture," says Fu Jing-hua, a retired researcher at the Chinese Academy of Chinese Medicine Sciences in Beijing and president of the Chinese Ancient Books Publishing House in Beijing. "Devoid of that cultural context, it would become a tree without roots."

Lofty ideals

But Zhang and Fang Shi-min, a US-trained biochemist who now runs a society called New Threads that is known for fighting pseudoscience and research misconduct in China say that it is exactly those traditional Chinese medicine theories that should be abolished. Conceits such as *yin* and *yang*, *wuxing* and *qi* "are inaccurate descriptions of the human body that verge on imagination," he says.

Inevitably, cultural factors may be the biggest obstacle in bridging the East–West gap. "The field of traditional Chinese medicine is notorious for being averse to criticism," says Yuan Zhong, a philosopher of medicine at the Chinese Academy of Medical Sciences. "If people are not allowed to disagree or voice their opinions, there would be

no hope of progress for any discipline."

But although heated exchanges are boiling over in debates on the future of traditional Chinese medicine, it's business as usual in Liu's practice. He is sanguine about the convergence between traditional Chinese medicine and Western medicine, but has a pragmatic attitude towards it. "Whether from the East or the West, we share the same goal of improving human health. As long as it works, anything goes," he remarks. But Liu says that he is yet to see any real progress in the merging of the two philosophies and, until then, his intuition and experience — as well as traditional Chinese medicine's seemingly arcane theory and practice — will serve him and his patients just fine. ■

Jane Qiu writes for *Nature* from Beijing.

"Whether from the East or the West, we share the same goal of improving human health."

— Liu Wen-long

1. Zhang, G.-Y. *Med. Phil.* **27**, 14–17 (2006).
2. Qiu, J. *Nature* **446**, 590–591 (2007).
3. Wang, M. *et al. Phytother. Res.* **19**, 173–182 (2005).
4. Wang, Y. *et al. Planta Med.* **70**, 250–255 (2004).

See Editorial, page 106.

Authors' financial interests should be made known to manuscript reviewers

SIR — Much emphasis has been put on the importance of policies that require authors to be transparent about financial conflicts of interest. *Nature*, for example, requires most authors to submit a declaration of any competing financial interests in relation to the work described in a submitted article. The reason why journals have these policies is, presumably, to safeguard the objectivity of the research. Transparency is thought to promote objectivity because if readers are aware of potential financial conflicts, they can critically evaluate the ways in which such interests may have affected the research — for example, in the selection of evidence, interpretation of results, or research methodology.

Yet transparency is insufficient as a safeguard of objectivity. Scientific expertise is necessary to correctly evaluate whether conflicts have biased the research, yet financial conflicts are revealed only when an article is published. This prevents peer-reviewers — who are in the best position to evaluate the possible influence of the conflicts of interest — from having access to the information. Thus, it is not clear to us how revealing financial interests in a statement accompanying publication of an article can allow readers to make accurate assessments of bias.

In addition, these policies foster an abrogation of scientific responsibility by the research community, because they put the burden of critical evaluation on the public, who in the main are not scientifically knowledgeable at a detailed level. This aspect is of particular concern for papers in journals such as *Nature*, which are likely to be widely disseminated to the public by the media. Even if biases are identified after publication and a correction is made, such criticisms tend not to be publicized to the same extent as the original article.

If the aim of conflict-of-interest policies is to promote objectivity and inform readers and the public, we believe a more effective approach would be for authors to be required to reveal possible financial competing interests, not only to the public after publication, but also to reviewers during the peer-review process.

Inmaculada de Melo-Martin*, Kristen Intemann†

*Division of Medical Ethics, Department of Public Health, Weill Cornell Medical College, 411 East 69th Street, New York, New York 10021, USA

†Department of History and Philosophy, Montana State University, 2-155 Wilson Hall, Bozeman, Montana 59717, USA

Comments on the suggestion made in this Correspondence, and on *Nature's* policy,

are welcome at Peer to Peer, the blog for and about peer review, at http://blogs.nature.com/peer-to-peer/2007/07/should_authors_financial_inter.html

Space-time safe — at least until the LHC switches on

SIR — You are either wildly optimistic, or confused, to suggest in your News Feature 'Extreme light' (*Nature* **446**, 16–18; 2007) that extreme laser experiments could "rip apart the fabric of space and time".

The phenomena of Unruh radiation and Schwinger pair production, which are the motivation for extreme laser experiments, simply probe the behaviour of quantum fields on a fixed, smooth space-time. Even if the experiments are totally successful, the only thing that could be 'ripped apart' is the vacuum state of some quantum fields.

If one considers quantum theories of gravity, there is no formal way to distinguish the space-time manifold from the quantum fields, because space-time itself should be quantized. (Don't ask me how.) But at low momenta and energy densities, relative to the fundamental scale of quantum gravity, usually around the Planck scale of 10^{19} GeV, such theories should reduce to something like the well-known picture of quantum field theory on a classical, weakly curved manifold described by general relativity, which has been successful in describing particle physics and optics.

Hence it is very difficult to probe quantum gravity experimentally, unless the low-energy limit of the theory gives some (necessarily small) deviation from this established picture. For example, varying 'constants', equivalence principle violations or Lorentz violation could amount to an indirect probe.

There are also special classes of theories in which the fundamental scale of quantum gravity may be much lower, even a few thousand GeV. However, these extra dimensions would not be probed by laser experiments, since photons do not propagate along them, unless the energy density approached the fundamental (TeV^4) scale. Whether an exawatt beam (rated by power, rather than energy density) could do so depends on its focusing and pulse length: I highly doubt it.

The main hope at present to probe theories with TeV fundamental scale comes from the Large Hadron Collider (LHC), which is designed to produce such large energy densities in some of its collisions.

It is conceivable that black holes might result from gravitational collapse of dense

regions — which would, finally, be a tale of ripping the fabric of space-time. Before the LHC switches on, space-time should be quite safe. Depending on your point of view, this may be a source either of relief or of disappointment.

Thomas Dent

Theoretical Physics, University of Heidelberg, Philosophenweg 16, 69120 Heidelberg, Germany

Mentors could support a student reviewer database

SIR — Your Editorial 'Mentors of the future' (*Nature* **447**, 754; 2007) highlights a need to encourage ethical, honest and fair peer review by young scientists. Although I applaud the ethos of the argument presented, graduate students such as myself often suffer from anonymity in their field of research, even though our work is often at the cutting edge. A lack of publications can render a student invisible to editors and may result in missed opportunities to offer their services as referees.

I suggest that journals consider introducing a 'PhD student peer-review pool' to which students and their supervisors can sign up. Such a database, including a student's name, area of research expertise and current supervisors, would provide editors with a ready supply of willing referees. Editors could try new referees in the knowledge that they will be supported during the review process by their supervisor, and could provide feedback to the student about the quality of the report.

Refereeing has often been described as a thankless task, but although it does require considerable effort, it also provides invaluable experience in critical interpretation of science. Having recently completed my first review, I believe that it has made me far more objective about my own writing and can only benefit the production of my thesis. I therefore look forward to receiving my next invitation to review.

Angelo P. Pernetta

Centre for Ecology and Hydrology, FBA East Stoke, Wareham, Dorset BH20 6BB, UK

Mentors: public lists would help students choose

SIR — Thank you for your excellent Feature, 'Nature's guide for mentors' (*Nature* **447**, 791–797; 2007). It's definitely one of the best things I've read in *Nature* in the 25 years I've been reading the journal. To better help students make informed choices about choosing a mentor, it would be enormously useful if public granting agencies such as the US National Institutes of Health would

publicly post the 'trainees' lists that are included in training grants for every faculty member in a given PhD training programme. If this were done, students considering applying to those labs would know their actual chance of being mentored successfully.

Ben Barres

Department of Neurobiology, Stanford University School of Medicine, 299 Campus Drive, Stanford, California 94305-5125, USA

How incompatibilities may have led to eukaryotic cell

SIR — In the Connections Essay 'Disappearing act' (*Nature* **446**, 983; 2007), James A. Lake discusses the evolution of the eukaryotic cell from the perspective of the origin of the eukaryotic gene content. As he points out, there are two main gene classes in prokaryotes (archaea and bacteria): operational genes, for "day-to-day processes of cell maintenance"; and informational genes, that "convert information from DNA into proteins". Eukaryotes are derived from archaea and bacteria, but curiously, archaeal operational genes and bacterial informational genes are almost completely absent from the eukaryotic genome. Lake suggests that "because two types of ribosomal genes cannot exist in the same nucleus, the archaeobacterial ribosome may simply have been the lucky survivor when one of the components in the eubacterial ribosome was inactivated." But, Lake asks his readers, why were the archaeobacterial operational genes eliminated?

One hypothesis is that the archaeal and bacterial cells that merged to form the eukaryotic cell were 'metabolically incompatible', and consequently their merger resulted in the elimination of the archaeal operational genes. But what, then, was the basis for this 'metabolic incompatibility' and for the bacterial genes to finally get the upper hand? David Valentine has suggested that adaptations to energy stress dictate the ecology and evolution of archaea (*Nature Rev. Microbiol.* **5**, 316–323; 2007). The biochemical mechanisms enabling archaea to cope with chronic energy stress include structural (a less ion-permeable membrane) and metabolic (pathways highly adapted to niches with low energy availability) components.

We suggest that during eukaryogenesis, the bacterial endosymbiont, which gradually became the mitochondrion, ended the chronic energy stress in the proto-eukaryote. In contrast to the anaerobic archaeal host, the aerobic bacterial endosymbiont was able to maximize the availability of energy and gradually become an efficient energy-converting organelle, the 'powerhouse' of the eukaryotic cell. The proto-eukaryotic

metabolism had to be reorganized accordingly. The unique archaeal adaptations to chronic energy stress were no longer advantageous, and were out-competed by the 'higher energy'-adapted bacterial metabolism. Lateral gene transfer from organisms other than the two founders certainly contributed to the eukaryotic gene repertoire, but we think that operational genes mostly originated from bacterial donors because the archaeal genes were incompatible with the energy-rich environment of the eukaryotic cell.

To return to Lake's Essay, the two faces of the eukaryotic gene content (the Janus paradox) might reflect two types of incompatibilities between its two prokaryotic founders: structural incompatibility between the informational systems; and environmental (or ecological) incompatibility between the metabolic systems. These eventually led to the synergetic, two-faced, chimaeric result — the eukaryotic cell.

Yaacov Davidov*, **Edouard Jurkevitch†**

*Department of Biological Chemistry, The Weizmann Institute of Science, 76100 Rehovot, Israel

†Department of Plant Pathology and Microbiology, The Hebrew University of Jerusalem, 76100 Rehovot, Israel

Text mining: powering the database revolution

SIR — Mark Gerstein and colleagues in Correspondence (*Nature* **447**, 142; 2007) propose that journals should require authors to manually provide structured abstracts to facilitate text mining of biological information. There are three main difficulties in implementing such a proposal.

First, life-science terminologies are huge, diversified and complex. This means that identifying the correct content descriptors is almost impossible for inexperienced users of online term repositories. For example, Medical Subject Headings (www.nlm.nih.gov/mesh), the International Classification of Diseases (www.who.int/classifications/icd/en) and Gene Ontology (www.geneontology.org) are high-volume — tens of thousands of terms — and structurally complicated terminological systems, each with different design rationales, naming conventions and principles of structural organization. Even human indexers, search specialists and database curators with routine exposure to these resources have to invest much effort in understanding and keeping track of their content as well as terminological updates and revisions. Will scientists find the time to dive so deeply into this alien terminological territory, and be capable of finding exactly what they are looking for?

Second, the coverage of existing

terminologies for the many subdomains in the life sciences is incomplete. The two main terminological umbrella systems for the life sciences, the Unified Medical Language System (<http://umlsinfo.nlm.nih.gov>) and the Open Biomedical Ontologies (<http://obo.sourceforge.net/main.html>), contain impressive numbers of individual terminologies, but their coverage of the life sciences is still fragmentary and suffers from varying depths of description. The size of the terminology gap is likely to be even more pronounced if authors were required to encode relational descriptions, for example indicating a binding relation between two specific proteins, P1 and P2, by Bind(P1, P2), because such a vocabulary has not yet been determined.

Third, the quality and reliability of author-supplied content descriptions is quite a hurdle. Even if the first and second problems were to be solved, human indexers, even professional ones, are liable to error as well as to the possibility of intrinsic subjective bias (M. E. Funk and C. A. Reid *Bull. Med. Libr. Assoc.* **71**, 176–183; 1983). This is not to say that authors of a structured abstract would consciously cheat, but rather there is a grey area of overstatement and overestimation of one's own results in a highly competitive scientific environment. If authors' structured entries were subject to peer review together with the submitted article, this would be more work for the reviewers as well as the authors — neither of them likely to have been trained as terminologists.

As an alternative, we suggest automated procedures for knowledge capture in which neither the authors nor the reviewers are in the loop. There has been significant progress in automatic text mining and information extraction as well as in the methodological foundations of life-science terminologies in terms of ontologies, knowledge representation languages and semantic encoding standards. These efforts in automating the generation of content descriptions and linking them directly to biological databases are strongly experimentally founded and would help to avoid additional workload and subjectivity — see, for example, the BioCreAtIvE competition results (<http://biocreative.sourceforge.net>). Once automated mechanisms for content analysis are applied, this also increases the coverage and the recency of the literature entered into biological databases, as human input is complemented by computationally generated content.

Udo Hahn*, **Joachim Wermter***, **Rainer Blaszczyk†**, **Peter A. Horn†**

*Friedrich Schiller University Jena, Computational Linguistics Group – JULIE Lab, 07743 Jena, Germany

†Institute for Transfusion Medicine, Hannover Medical School, 30625 Hannover, Germany

COMMENTARY

Beyond the troubled relationship

If scientific culture in the Muslim world has changed since the golden era of Islamic science, so has the practice of Islam. Reintroducing knowledge and creativity requires a revival of both, argues **Ziauddin Sardar**.

On Islam and science two things can be stated with some certainty. One, science thrived during the classical period of Islam; two, science in Muslim society has suffered a drastic decline. The difficulties arise in trying to ascertain when the decline began and what the causes were. Historians of science offer different dates and reasons.

It is tempting to blame Islam itself. There is something in the teachings of Islam, the argument goes, which does not allow science to take root in Muslim societies. This suggestion not only belies history but also the basic teachings of Islam, which proclaims itself as an intrinsically rational world view.

There are some 800 verses in the Qur'an that invite the reader to think and to examine the material world, using reason to understand nature. The sayings of the Prophet Muhammad reinforce these teachings, emphasizing that understanding comes through scientific endeavour. "An hour's study of nature is better than a year's prayer," the Prophet declared. He directed his followers to "listen to the words of the scientist and instil unto others the lessons of science". And the Prophet made an essential distinction: the Qur'an, as well as his own teachings, were an invitation to reason and study what exists and can be discovered — not scientific pronouncements in and of themselves.

If the basic teachings of Islam are the same now as they were 1,400 years ago, what was it that drove science, learning, knowledge and creativity from Muslim culture? Historians have tried in vain to fix a date, to pinpoint what provoked the downward spiral. There are many factors to consider — including colonialism and wars — but in searching for the genuine causes of decline, I believe we must consider that the practice, if not the teachings, of Islam has changed. By recognizing this, I will argue, Islam can then become part of the solution.

Gradual decline

As a faith and a culture, Islam is a holistic world view. Muslims are fond of describing Islam as a 'total way of life'. What they mean is that the world view of Islam integrates all aspects of reality through a moral perspective. This perspective is provided by a



The Qur'an invites readers to use reason to study nature, so what drove science from Muslim culture?

framework of conceptual values within which Muslims endeavour to answer human problems. Concepts such as *ilm* (knowledge), *ijma* (consensus) and *istislah* (public interest) are the driving forces of Muslim society. *Ilm*, the urge to know, and *ijtihad*, the quest for sustained reasoning, were central to classical Islam and produced a culture with science and rationality at its core.

The question today is how these driving forces sputtered to a halt. I would argue that the decline of science in Muslim societies is a product of the systematic reduction in the meaning of the basic concepts of Islam. This process not only reduced Islam from a holistic world view to a one-dimensional faith but also truncated the creativity of Muslim societies. If Muslim society is like a human body — an analogy once used by the Prophet Muhammad — then this process of reduction has taken the mind from the body. What is left is living and functioning — but without the brain. As a result, science and rationality have almost completely evaporated in contemporary Islamic culture.

There is a conventional explanation for this process. Studies endlessly refer to a time when 'the gates of *ijtihad*' were closed by a few

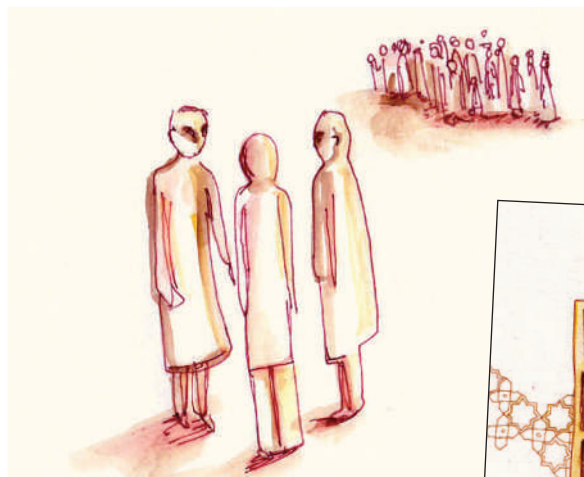
religious scholars. But the use of *ijtihad*, or sustained reasoning, was never actually banned: the shutters came down only slowly. And the dictate of the religious scholars has been disputed by every Muslim reform movement from that day to this.

Under colonialism

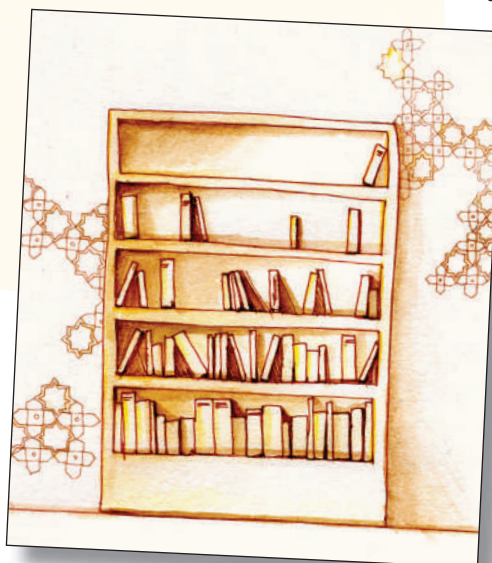
So what is missing from the standard analysis? I have argued for the intimate and constructive connection between Islam and science. But Islamic science not only had a conceptual religious context, it also had a cultural context. In history, as now, science was integral to the development of society. In Muslim history it fed the growth of industries — studies of mineralogy, advances in chemistry, botany and hydraulics all had applications relevant to a productive society. And agricultural advances, supported by the study of agronomy, hydrology and geology, raised living standards and created wealth.

Muslim civilization wherever it spread was a city-building culture; architecture, city planning and land management, provision of clean water and sewage disposal all benefited from the application of science. The study of astronomy, geography and map making, alongside development of compasses and observational instruments, all facilitated long-distance trade.

"This process reduced Islam from a holistic world view to a one-dimensional faith."



Knowledge and consensus, central concepts in Islam, have narrowed over time.



Consequently, the state was a major sponsor and consumer of science for multiple reasons — from building cities to warfare technology.

The great scientists of the classical era all contributed, directly or indirectly, to this social context. In turn, the institutions of Islam ensured that science prospered and served society. One of the pillars of Islam is *zakat*, the annual payment required of all Muslims and dedicated to social purposes such as education, health provision and the eradication of poverty. Charitable giving begat new social institutions called *waqf*. These charitable endowments funded projects such as hospitals, universities and research establishments. Science never emerges in a vacuum; it always has a cultural context; it is fed and shaped by the conditions of its time and place.

If the Muslim world had not been such a vibrant, dominant, growing concern in the fifteenth century, Europe would have had no need to subvert its power. What is termed the age of exploration was a deliberate strategy, diligently pursued by European nations to circumvent what they regarded as a Muslim stranglehold on their economies, a stranglehold that was maintained by Islamic science. Investment in exploration gave a new impetus to the development of science in Europe. It also led to colonialism.

Colonialism had two outcomes in Muslim civilization. First, it suppressed and displaced meaningful cultural activities from Muslim society. It did this by introducing new systems of administration, law, education and economy all of which were designed to instil dependence, compliance and subservience to the colonial powers. The decline of Islamic science is one aspect of the general economic and political deterioration of Muslim society that resulted.

In the colonial context, Islam as a holistic way of life became mere rhetoric. Islamic education became a one-way ticket to marginality. What relevance could such education have when its concepts and principles had no practical meaning for how society operated? Nor was the 'modern' education offered by the colonial powers a path to success: it was intended to produce subjects that served colonial needs.

Western education ruled, but it taught the colonized to accept a distorted version of their own history as backward, their own science as irrelevant, and their own medicine as nothing but mumbo jumbo. The dismissal of Islamic science as 'real' science went hand-in-hand with the dismissal of the entire conceptual order of Islam as religion, culture and civilization.

Second, Islam itself was reduced to a defensive enclave stubbornly holding on to the few remnants of authority left to indigenous control. Colonialism turned the pursuit of knowledge by Muslims from a broad cultural activity into a narrow religious one.

Narrowing of faith

In classical Islam, both the definition of knowledge and its classification was a major intellectual activity. An *alim* (scholar) was anyone who acquired *ilm*, or knowledge, which was itself described in a broad sense. Knowledge meant everything from science

and philosophy to art, literature and theology. We can see this in the early classifications of knowledge by such scholars as al-Kindi, al-Farabi, ibn Sina, al-Ghazzali and ibn Khaldun. So all learned men, scientists as well as philosophers, scholars as well as theologians, constituted the *ulama*.

Under colonialism, Muslim society revisited the question: who is a scholar and what makes him an authority? Whose thought, research and opinion were worthy of social and cultural attention? And as *ijtihad* became increasingly irrelevant, *ilm* was increasingly reduced to religious knowledge and the *ulama* came to constitute only religious scholars. From an Islamic point of view, only the pursuit of religious knowledge came to be seen as important.

Similarly, the idea of *ijma*, the central notion of communal life in Islam, has been reduced to the consensus of a select few. *Ijma* literally means consensus of the people. The concept dates back to the practice of Prophet Muhammad. When the Prophet wanted to reach a decision, he would call the whole Muslim community — then admittedly not very large — to the mosque. A discussion would ensue and the entire gathering would reach a consensus. Thus, a democratic spirit was central to political life in early Islam. This disappeared under colonialism.

The notion of *ijma* had another function. It located authority in argument. If *ilm* was intended to spread knowledge throughout society then most citizens would be members of a learned community. The consensus of the community would therefore be a learned one arrived at on the basis of argument. But the reduction of *ilm* removed knowledgeable citizens from the equation — *ijma* was truncated to 'the consensus of religious scholars'. Thus both democracy and the pursuit of knowledge as a whole became irrelevant to Muslim culture.

Other key elements of the conceptual framework of Islam were similarly narrowed. *Jihad* ceased to be an intellectual struggle and was abridged to warfare. *Istislah*, or 'public interest', a major source of Islamic law and an important



Charitable endowments, one of the pillars of Islam, helped shape science.

impulse for science in the classical period, all but disappeared. And *ijtihad* came to mean little more than a pious desire.

Thus, the question of how Muslim societies can rediscover the spirit of scientific inquiry is not just a question of policy, funding or building prestigious institutions in the name of progress. Mostly, it is about how the enterprise of science is made relevant and meaningful within Muslim culture. The decline of Islamic science was a product of combined forces that engineered a conceptual reduction in Muslim civilization.

Sacred science

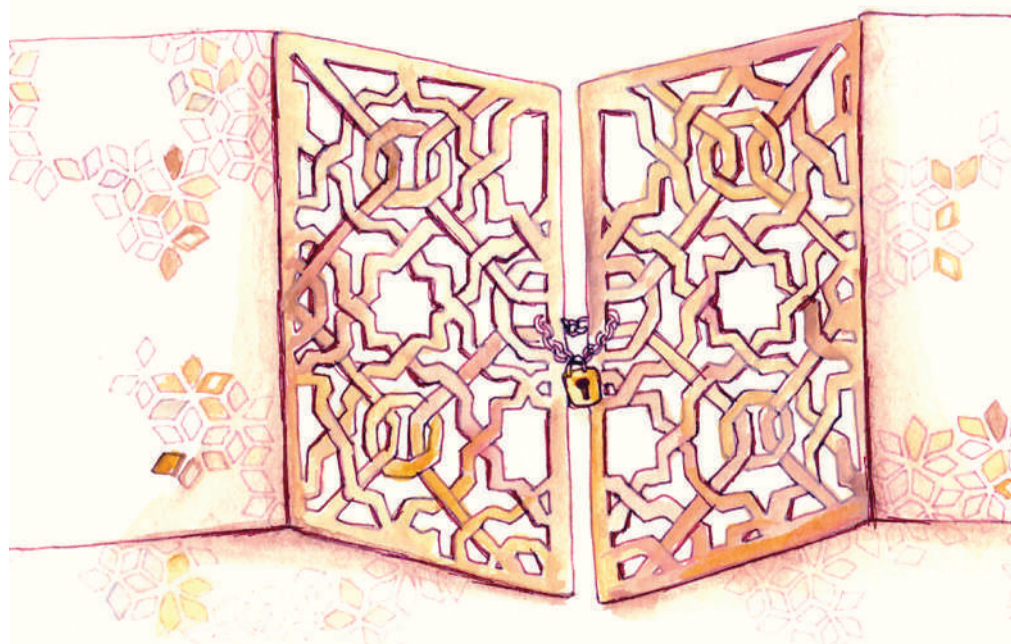
My proposition is clear: the decline can only be reversed by a conceptual shift. Science will only take root in Muslim societies if they can rethink Islam itself as a holistic enterprise. Science will flourish, paradoxical as this may seem to many, when Islam re-emerges as an integrative way of knowing, being and doing; when it reconstructs the open intellectual climate and cultural paradigms it once sustained.

When in 1980, I initiated a discussion on Islamic science, first in the pages of *Nature* and then *New Scientist*, I was not fully aware of the problem with these reductive concepts in Muslim societies. By 'science' I meant and understood an objective and systematic endeavour within the ethical framework of Islam, motivated by Islamic injunctions on the virtues of reason and the pursuit of knowledge. I turned out to be totally wrong.

If we look at what is meant by 'Islamic science' today, there are two trends. The first derives from the fundamentalist idea that all knowledge can be found in the pages of the Qur'an. This is a further reduction of the concept of *ilm*. A lavishly funded project in Saudi Arabia — Scientific Miracles in the Qur'an — has generated books, papers and journals investigating the scientific content of the Qur'an. Relativity, quantum mechanics, Big Bang theory, embryology and much of modern geology has now been 'discovered' in the Qur'an.

Meanwhile, 'scientific' experiments have been devised to discover what is described in the Qur'an but not known to science. For example, a project to harness the 'energy of the *jinn*s' enjoyed much support in the mid-1990s in Pakistan. This reductive fundamentalism now embraces creationism and is generating a growing movement for 'intelligent design' in the Muslim world.

The second trend is best described as mystical fundamentalism. In this perspective, the consensus, the *ijma* of a knowledgeable community, is reduced to a group of mystics with secret and sacred scientific knowledge. Islamic science is no longer a problem-solving enterprise; it is a mystical quest for understanding the absolute. In this Universe, conjecture and hypothesis have no place; all inquiry must be subordinate to mystical experience. The mate-



Can we reopen the 'gates of *ijtihad*' and revive the quest for reason that was once central to Islam?

rial Universe is studied as an integral part of higher levels of existence and consciousness.

Even the history of Islamic science has been rewritten to emphasize the occult, alchemy and astrology — at the expense of the vast amount of work in exact sciences — in an attempt to show that Islamic science was largely 'sacred science'. In many academic circles, this mystical tendency has acquired a strong presence.

Old way of thinking

These two trends, the fundamentalist and the mystical, suggest that real science has almost completely disappeared from Muslim consciousness. In an article last year, (*Nature* 444, 22–25; 2006) Ehsan Masood wrote "today's Muslim states barely register on indices of research spending, patents and publications". And he concludes the situation is not just bad; it is set to get worse.

However, I don't think we need to be so pessimistic. The solution to any problem begins with a diagnosis; this diagnosis has already begun. The realization is growing that science is important not just for the prosperity of Muslim societies, for economic development, for misplaced political vanity or for acquiring nuclear weapons — but that it matters because it is vital for the recovery and survival of Islam itself.

This is the main message of the 2003 Arab Human Development Report on 'Building a Knowledge Society'. It admits frankly that Muslims cannot merely continue to blame everything on colonialism and the West. Muslim states have failed, by their own Islamic standards, the challenge of independence. The report blames authoritarian thought, lack

of autonomy in universities, the sorry state of libraries and laboratories, and under-funding in the Arab world.

Moreover, the report recognizes the conceptual challenges, calling, in particular, for 'reviving *ijtihad*' as the driving force for change. Indeed, it is now widely argued that science can play an important part not just in re-establishing *ijtihad* but in making Islam whole again, reuniting reason with revelation. This is the course that the Organization of the Islamic Conference — a 57-member intergovernmental organization of Muslim states — has set itself.

Signs of change can even be detected in religious institutions. The Al-Azhar University in Cairo, one of the most influential institutions of the Muslim world, for example, has now opened up to science. In colonial times, it abandoned science to become a religious institution concentrating on theology. Now it is rebalancing its curricula, emphasizing science as much as religion. Religious classes are making way for laboratories and science courses. Similar developments are taking place in Indonesia and Turkey.

There is no doubt that Muslims have a deep emotional attachment to their scientific heritage. But contemporary Muslim society needs more than nostalgic pride in a long-departed golden age. It has to reinstate the way of thinking, the critical consciousness and methodologies that made Islamic science possible; and it must make this way of thinking and knowing relevant to contemporary times. There is every reason to hope that the revival of science and a reform agenda for Islam in Muslim society will proceed hand in hand. ■

This article is an edited version of a lecture that Ziauddin Sardar, a London-based writer, first gave to the Royal Society in London.

"Science never emerges in a vacuum; it always has a cultural context."

BOOKS & ARTS

Imagine a better world

Given that humans are here to stay on Earth for some time yet, what can we do to lessen our impact?

The World Without Us

by Alan Weisman

Thomas Dunne Books/Virgin Books: 2007.
288 pp, \$24.95/£20.00

Stuart L. Pimm

Imagining “the world without us” is such a clever idea. It powerfully motivates thinking about what the 6 billion of us do to Earth. It conjures, too, reflections on the sometimes surprising parts of Earth that we have left alone. How these have fared might shape future environmental choices.

The examples in Alan Weisman's *The World Without Us* are familiar from numerous other books about the planet today. Of course, good science books assemble what is already known to the academic priesthood to make the ideas widely accessible. They strongly influence that priesthood, embarrassed as we often are to admit we found the original literature unintelligible. Great books synthesize facts and ideas onto a sweeping canvas. Sadly, Weisman's book does not assemble, synthesize, step back from its details to pose larger questions, engage the reader in a debate about the salient features of human impact or present solutions. It has no introductory and final chapters laying out key issues.

The book does come frustratingly close to being more than its collection of set pieces. Chapter one has a photo of the massive trees in Bialoweza forest, Poland. Many trees get that way — if we allow them to. Without us, one can almost imagine the loud sucking sound of Earth's trees growing to old age, taking back the carbon. Without us, some 15 million square kilometres of former cropland and another 5 million of tropical forest cleared for grazing would re-grow, sucking up roughly 10 gigatonnes of carbon each year for the next 30 years or more. Even ignoring other sinks, including soils, this very rough calculation suggests an immediate, significant reduction of atmospheric carbon, with attendant climate changes. The all too obvious practical answer to Richard Branson's challenge to remove carbon dioxide from the atmosphere is growing trees. Without us, there are lots of places for them to re-grow; even with us, there are too.

Without us, the nitrogen junkies of the global flora would protest the halving of their annual fix. Human chemistry has doubled the availability of fixed nitrogen through



There are still places that exist without us, such as the Bialoweza forest in Poland.

fertilizer in the biosphere. Without us, trees would flourish, nitrogen-dependent weeds would not, and most domesticated crop plants would die within the year.

Weisman worries about the bison in Bialoweza, imprisoned there by its fence. He moves on promptly to discuss how our buildings would decay and how the subways of New York would flood if we were gone, a subject of a morbid fascination, no doubt, to those who enter them daily.

He misses an opportunity in the chapter on “the world before us”. Our nearest relatives remain in this book's conceit. Evolution might not recreate us, but what about an ape with our intelligence? Palaeoanthropologists argue that man is just the forest ape who moved into the savannas in compelling (if often contradictory) creation stories. So, what is to stop a repeat performance? Weisman concludes that the aggressive common chimpanzee (*Pan troglodytes*) will overwhelm the meek — and much rarer — bonobo (*Pan paniscus*). Would this prevent a new ‘man’ from following in our path, or is an aggressive *Pan* the key step? This is an interesting, but unexplored possibility.

Such debates are fun over morning coffee. Other issues are practical and vital. There are places without us and we could designate more. Some of them are in this book — the demilitarized zone between the two Koreas, the Bialoweza forest, a few isolated reefs, decreasing areas of rainforest. These examples broach the biggest challenges of this century: can we reverse global carbon emissions, cool off the planet and prevent the irreversible loss of biodiversity?

Weisman discusses how nature returns in our absence. Allowing native trees to recover — the Lorax solution — so that both the trees and their friends will come back is also a good start to solving the biodiversity problem. He also discusses the constraints we have placed on recovery. For example, we have introduced weeds — that is, species that belong in other places. But we have not followed US environmentalist Aldo Leopold's first law of intelligent tinkering: to keep every cog and wheel. Nor have we followed the second law: to keep the instructions. There is no guarantee that even with all the pieces, we would be able to put nature back together again. The book

J. KORBEL

successfully warns us that almost everywhere are human artefacts that will be with us for a long time — plastics, for example — making things ever more complex.

These and many other good leads are in *The World Without Us*, but their placing is incoherent. There is not going to be an Earth without us. So excluding us from some of it and reducing our impacts elsewhere is an important part

of a discussion that this book could do a lot to facilitate. Perhaps the broader vision — the missing introductory and final chapters it so clearly needs — could be added on a related website.

Stuart L. Pimm is professor of conservation ecology at the Nicholas School of the Environment and Earth Sciences, Duke University, Durham, NC 27708, USA.

Degrees of change

Six Degrees: Our Future on a Hotter Planet

by Mark Lynas

Fourth Estate: 2007, 384pp, £12.99

Stefan Rahmstorf

This book is not for the faint-hearted. British writer Mark Lynas ventures where few scientists would dare to tread. He sets out to answer the question that many of us climatologists ponder in private and often get asked by journalists, but usually shy away from answering: what will it really be like to live on a warmer planet? In Lynas's words: "Will we all, as some environmentalists suggest, be reduced to eking out a living from shattered remains of civilization in Arctic refuges, or will life go on much as before — if only a little warmer?"

Lynas sets out to answer this systematically on the basis of his extensive reading of the scientific literature. He has spent many months in Oxford University's Radcliffe Science Library trawling through thousands of papers. The result is arranged in six chapters, one for every degree Celsius of potential global warming. His statements are referenced throughout, and, as a palaeoclimatologist, I was familiar with fewer than half of the 500 or so papers he cites. That is the nature of scientific specialization: few researchers could afford the time for such a wide-ranging literature review.

One of the best aspects of *Six Degrees* is that it pulls together data from past climate changes in Earth's history to get an idea of what a warmer climate might look like. The often dramatic natural climate changes of the past are sometimes cited by those opposed to reducing greenhouse-gas emissions as evidence of why we need not worry about the ongoing warming. But the latest Intergovernmental Panel on Climate Change (IPCC) report (www.ipcc.ch), which contains a 65-page chapter reviewing the main palaeoclimatic findings, shows that the first lesson from the past is that Earth's climate system is very sensitive. It has always responded strongly to natural changes in Earth's energy budget — the balance between the absorbed incoming energy from the Sun and the outgoing long-wave (thermal) radiation from Earth. So we have every reason to believe that it will respond strongly again to the growing

perturbation to the radiation balance that humans are now causing. *Six Degrees* hammers home the second lesson to be drawn from the past: mega-droughts or sea-level changes of tens of metres accompanied past cooling and warming episodes of similar magnitude (albeit probably not as rapid) as that expected this century.

I have my quibbles with some of Lynas's interpretations and there is the odd error, but such complaints seem petty in view of the overall achievement and importance of this book. Lynas avoids the obvious pitfall of cherry-picking the most dramatic possibilities; *Six Degrees* is alarming but not alarmist. He tries to give a level-headed account of what we may expect, mentioning scientific controversies where they exist. For example, he gives balanced discussions about the future of the Sahel region of northern Africa, citing studies that conclude that global warming may end the ongoing drought there, and about the risk of a shut-down of the North Atlantic Current, to which the new IPCC report assigns a probability of up to 10% this century.

Six Degrees is essentially about risks, because much remains uncertain about the future. My major beef is that he often makes risks sound like truths. After a sensible discussion of amplifying feedbacks from the carbon cycle and methane release, in which he states how uncertain and hard to quantify these still are, he then claims that three degrees Celsius of warming "inexorably leads to four degrees, which leads inexorably to five". Even a small risk of this happening is bad enough, without making it sound so definite.

Lynas, who is not a natural scientist, must be highly commended for basing his book thoroughly on science — more so, in fact, than some popular books on climate change written by scientists. Gloomy as his story sounds, in some cases he may even be too optimistic. The possibility of violent conflict in regions struck by drought and food shortages is only mentioned in his three-degree-warming chapter. A recent report by the German Advisory Council on Global Change (www.wbgu.de/wbgu_jg2007_engl.html) sees this risk arising much sooner.

The book ends with a lucid description of the state of denial about climate change that humanity is still in (but hopefully now emerging from), and a good account of the policy options available to us to stop global warming. Lynas is interested in leaving the reader ready for action, rather than depressed.

Lynas is a gripping story-teller, making the book infinitely less tedious than the papers it is based on. A must-read for those who can stomach it.

Stefan Rahmstorf is a climatologist at the Potsdam Institute for Climate Impact Research, Telegrafenberg, 14412 Potsdam, Germany.



M. HARVEY/NHPA

Climate change will bring water shortages to many regions, but it may end ongoing drought in others.

Culture costs lives

Vaccine: The Controversial Story of Medicine's Greatest Lifesaver

by Arthur Allen

Norton: 2007. 512 pp. £17.99/\$27.95

Michael B. A. Oldstone

Vaccines, like antibiotics and improvements to sanitation and water quality, have greatly affected human health. Vaccines play a commanding role in controlling such formerly devastating plagues as smallpox, yellow fever, measles, poliomyelitis, tetanus and diphtheria. The first vaccine to reduce the risk of developing liver cancer (hepatitis B) is now a reality. Nevertheless, some continue to regard vaccination as controversial, as Arthur Allen describes in his book on the evolution of vaccines, associated fears and dissatisfactions as well as related politics and players.

Allen is an investigative reporter who learned, just before his son was to receive a whooping-cough vaccination, that a new, 'safe' vaccine was replacing an older 'less safe' one. He then began researching the controversy over vaccines and the question of mandatory vaccination for those attending state-funded schools in the United States. He analyses why about 25% of US parents are reluctant to have children vaccinated, the social history of vaccination and the reasons for public ambivalence. Allen also deals with use of the Internet to obtain medical information that is frequently subjective, emotional and erroneous.

Routine immunizations to prevent crippling polio or measles are carried out in developed countries, such as the United States. The rationale for such mass vaccination is that if sufficient numbers of a population are vaccinated, the incidence of disease will be low or nonexistent due to herd immunity. As Allen notes, those in the basic sciences or medicine consider such issues objectively — that is, is the benefit of preventing disease greater than the risk to those vaccinated? Other opinions often reflect the political divide in the United States between those who believe in a free market and individual choice, and those who favour collective responsibility and government financing of programmes to protect the health of all children.

Allen's early chapters on the origins of vaccination are adequate, although his description of medical practices in the eighteenth and nineteenth centuries seems excessive, considering that the abysmal usage of purges, poisons and bleeding before normal physiology and events causing disease were understood. That bacteria or their toxins cause disease has only been known for just over 160 years, viruses were first isolated



Raggedy Ann became the symbol of a vaccine-damaged child.

little more than 100 years ago and the emergence of initial concepts about immunology and allergy followed. The history of Edward Jenner and the first vaccine (smallpox) along with the difficulty of its contamination with diverse infectious agents are well detailed. Interestingly, the religious objection dating back to Jenner's time in the late 1700s that man must not, by using vaccines, interfere with God's will has not abated. Similar arguments, and others, are raised today against using the vaccine for the sexually transmitted human papilloma virus — to prevent cervical cancer — in the United States and the polio vaccine in Nigeria.

Some factual errors mar certain chapters, and others are difficult to follow. For example, the chapter on measles also covers diphtheria, pertussis (whooping cough), tetanus, SV40 (simian virus), HIV/AIDS, varicella (chicken pox), hepatitis B, *Streptococcus*, respiratory syncytial virus, cowpox and the flu epidemic of 1919, a mix sufficient to disturb any reductionist.

Nevertheless, the book has several gems that will intrigue readers. One is the origin

of Raggedy Ann, a floppy cloth doll with hair of red yarn that inspired the children's books. It was made for Marcella, the daughter of New York City illustrator Johnny Gruelle, after she fell ill following smallpox vaccination. Raggedy Ann, with her limp limbs, became the symbol of a vaccine-damaged child. Another good anecdote follows the story of Norman McAllister Gregg, an Australian ophthalmologist who in 1941 linked rubella (German measles) infection during the first trimester of pregnancy with birth defects. About 20 years later, almost 50,000 women who were not protected by vaccination became infected with German measles after which 67% of their children had birth abnormalities, many severe. One alternative was illegal abortion, and that discovery became one of the catalysts in legalizing abortion.

The chapter on vaccines and autism presents the conservatism of US Congressmen Dan Burton (Republican, Indiana) and Curt Weldon (Republican, Pennsylvania), their attack on the National Immunization Program and their work with antivaccination groups. They argue that legislatures, as representatives of the people, and not scientists or public-health workers should decide which vaccines to administer to children. The proposition leaves us, to quote Allen, "...punched in the nose by a congressional bully. It does not matter that they were smart kids (scientists, public health officials), he (the bully) still controls the playground." In the words of the late Democratic senator for New York, Daniel Patrick Moynihan: "Every man is entitled to his own opinion, but not to his own facts."

What becomes clear throughout this book is that, when facts tangle with culture, culture often wins. For those interested in the politics and debate of compulsory vaccination, and the personalities involved in all sides of the fight, *Vaccine* is a good read. ■

Michael B. A. Oldstone is professor and head of the Viral-Immunobiology Laboratory in the Molecular and Integrative Neurosciences Department, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, USA.

S. SKOLD/ALAMY



BRITISH MUS., LONDON/BRIDGEMAN ART LIBRARY

Linnaeus and taxonomy in Japan

Doctors at the Dutch Trading House on Dejima were a conduit for science into and out of Europe.

His Majesty The Emperor of Japan

In memory of Carl Linnaeus I would like to address the question of how European scholarship has developed in Japan, touching upon the work of people such as Carl Peter Thunberg, Linnaeus's disciple who stayed in Japan for a year as a doctor for the Dutch Trading House (pictured above) and later published *Flora Japonica*.

In the first edition of *Species Plantarum* in 1753, and in his later books, Linnaeus described many Japanese plants and gave them scientific names. *Camellia japonica*, for example, was described in the first edition of *Species Plantarum*, and this scientific name is still used today.

These Japanese plants were illustrated by Engelbert Kaempfer in his book *Amoenitatum Exoticarum*, published in 1712. Kaempfer was a German doctor who served in the Dutch Trading House in Japan for two years from 1690.

At that time, Japan had isolated itself from the world. Japanese people were not allowed to go abroad, and visits by foreigners to Japan were severely restricted. As the policy of isolation was taken to suppress Christianity, the Dutch, who came for trading purposes only, were permitted to come to Japan.

Dutch people were made to live on an artificial island, Dejima, built in the sea off Nagasaki and connected to land by a bridge; they could not leave the island without permission. The head of the Trading House, however, was to visit the shogun at Edo — present-day Tokyo — once a year,

accompanied by his delegation including the doctor. Kaempfer thus visited Edo twice, taking more than 80 days for the trip each time. The 256 sketches of Japanese plants he made during his stay are now kept in the Natural History Museum, London.

Rates of exchange

In 1775, 83 years after Kaempfer left Japan, a Swedish doctor, Carl Peter Thunberg, arrived at the Dutch Trading House. Thunberg was Linnaeus's pupil and later became a full professor at Uppsala University in both botany and medicine.

Whereas Kaempfer had just one visit from a Japanese doctor during his two trips to the city, Thunberg was visited by five doctors and two astronomers the moment he arrived in Edo. Japanese doctors had developed a deeper recognition of European medicine.

This change occurred because in 1720, Shogun Tokugawa Yoshimune relaxed the prohibition on importing books, which had been put in place to prevent Christian ideas from coming into Japan. This development stimulated research on European science, and people came to focus their attention on medical books written in Dutch. In 1774, for example, *A New Book of Anatomy* was published. This Japanese translation of a Dutch book was made by doctors of Edo, including Sugita Genpaku, Katsuragawa Hoshu, a physician for the Shogun, and his friend Nakagawa Jun-an, after they saw a dissection of a human body and were convinced of the accuracy of the Dutch book.

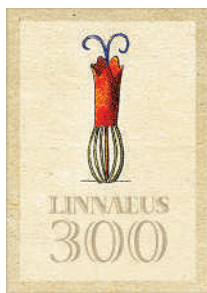
Katsuragawa Hoshu and Nakagawa Jun-an visited Thunberg almost every day and sometimes stayed until very late into the night to learn from him about various scientific matters. Both of them, Nakagawa Jun-an in particular, could speak Dutch quite well. In his book *Travels in Europe, Asia and Africa Made During the Years 1770–1779*, Thunberg writes that he asked Katsuragawa Hoshu and Nakagawa Jun-an the Japanese names of the fresh plants that they brought, and taught them the Latin names and the Dutch names of the plants.

Exchanges between Thunberg and the two Japanese doctors continued even after Thunberg's return to Sweden. The letters are kept in Uppsala University. I saw them with Their Majesties the King and Queen of Sweden during our visit in 1985, as Crown Prince and Crown Princess, and it left a deep impression on both of us.

End of an era

We do not know exactly when the scientific names under the binomial nomenclature, originated by Linnaeus, were introduced to Japan. Despite what Thunberg writes in his book, some doubts remain that they were first introduced to Japan at his time. They started to be used here after a German doctor, Philipp Franz von Siebold, arrived at the Dutch Trading House in 1823. By this time, there were many Japanese who could speak Dutch. Siebold established a school of medicine and a clinic for treating patients in the suburbs of Nagasaki. He could also leave the island of Dejima to visit patients at their homes or to collect medicinal herbs.

In 1829, Ito Keisuke wrote a book in



which Linnaeus's nomenclature was used for the first time in Japan. Keisuke took the scientific names of plants in Thunberg's *Flora Japonica*, which Siebold had brought to Japan, put them in alphabetical order, and added their Japanese names. In his book, Keisuke introduced Linnaeus's classification system as the "Explanation of the 24 Classes".

Keisuke studied under Siebold for six months in Nagasaki. When he was about to return home to Nagoya, he was given Thunberg's book as a gift. Keisuke sent the manuscript of his book, *A Translation of Thunberg's Flora Japonica*, to Siebold in Nagasaki, and Siebold checked it.

In 1854, Japan and the United States signed the Treaty of Peace and Amity as the arrival of the American naval fleet brought to an end Japan's policy of isolation, which had lasted for more than 200 years. After that, Japan began establishing diplomatic relations with many countries. The last shogun, Tokugawa Yoshinobu, resigned from his post in 1867, and a new government was formed under Emperor Meiji.

The Meiji government sent students overseas and invited foreign teachers to Japan, and the Japanese people made a great effort to acquire Western knowledge. The foreign teachers who were invited to Japan at this time made a great contribution to Japan, and the students who went to study overseas also contributed in various ways to the subsequent development of Japan.

New shoots

One of the academic achievements made by Japanese scientists in the nineteenth century was the discovery of gymnosperm sperm. In 1896 Hirase Sakugoro, an illustrator in the botanical laboratory of the University of Tokyo who later became a research associate, observed the swimming of ginkgo sperm, and published his paper on this discovery in the *Botanical Magazine* (Tokyo).

A month later, Ikeno Sei-ichiro, an associate professor in the agricultural department of the University of Tokyo who collaborated with Hirase Sakugoro, found cycad sperm, also reported in the *Botanical Magazine* (Tokyo).

This discovery was not believed at first, but it became accepted after *Zamia* sperm, from the same cycad family, was discovered in the United States the

following year in 1897. For this achievement these two researchers were awarded the Imperial Award of the Japan Academy in 1912.

The ginkgo is unique in its phylogeny because it is a single-order, single-family, single-genus, single-species plant. It flourished in the Mesozoic Jurassic age but survived only in China, and was brought from China to Japan in ancient times. It was given a scientific name by Linnaeus, on the basis of Kaempfer's illustration.

The ginkgo tree that Hirase Sakugoro used for his research is still standing in the Koishikawa Botanical Gardens of the University of Tokyo. I visited with the Empress last year and looked at the ginkgo tree, thinking of the research that was done a long time ago.

In the twentieth century, as Japanese taxonomy made progress, more and more new species

began to be reported. Before that, Japanese animals and plants were given scientific names by European scientists, and as a matter of course, the type specimens used for naming them were kept in European museums. Therefore, when Japanese researchers wanted to describe a Japanese animal or plant as a new species, they had to check the type specimens in foreign countries one by one, and the difficulties they encountered were far from trifling.

Thanks to the efforts made by many people, all Japanese spermatophytes, pteridophytes and vertebrates excluding fishes now have scientific names. However, there are still many unnamed fishes; in particular, there are many gobioids that must be given scientific names.



Linnaeus referred to Kaempfer's drawings.

When I started my research, I frequently referred to a book titled *Fish Morphology and Hierarchy* by Kiyomatsu Matsubara, published in 1955. The book covered all Japanese fishes with keys to the species, and it listed 134 gobioids, including subspecies. In the more recent *Fishes of Japan with Pictorial Keys to the Species*, published in 2002, the number of gobioids, including subspecies, increased to 412, but 45 of them have only Japanese names and have no scientific names yet.

Class act

In 1967, I published a paper in the *Japanese Journal of Ichthyology* on the classification of the four gobioid species of the genus *Eleotris* found in Japan based on the arrangement of their sensory papillae. In those days, no one in Japan was yet classifying gobioids thus, and some people had considerable doubts about my classification. The arrangement of the sensory papillae has now become an important factor in classifying gobioids, and I am glad that I have been able to make some contribution in this field.

The binomial nomenclature established by Linnaeus has been immensely beneficial, providing a universal basis for taxonomy and enabling taxonomists around the world to communicate with each other through a common language about things existing in nature.

Today, an even newer field of research, molecular biology based on evolution, is seeing remarkable development. As a result, systems based on phylogeny are considered to be more accurate and are now the mainstream of taxonomy.

As I have been familiar with classifications based on morphology since I was young, the appearance of the electron microscope, which enabled me to observe minute morphological characteristics, and my encounter with an even smaller world, where classification is based on DNA analysis at a molecular level, have been great experiences for me as a researcher.

On the 300th anniversary of Linnaeus's birth, I feel that taxonomy is entering a new era. The analysis of mitochondrial DNA will open up great possibilities of discovering new species that cannot be distinguished morphologically. I hope to understand and take into consideration this newly developing field of research, but at the same time, I intend to continue to give my attention to and keep up my interest in morphology, which is a field of study carried on from Linnaeus's days.

This is an edited extract of a speech given by His Majesty The Emperor of Japan to the Linnaean Society of London on 29 May 2007. The full text of this speech (<http://tinyurl.com/29djvx>) will be published later this year by *The Linnaean*.



Carl Peter Thunberg, pupil of Linnaeus.

KVA

NATURAL HISTORY MUS., LONDON

NEWS & VIEWS

PHOTONICS

Light in chains

Niek F. van Hulst

Diffraction places a fundamental limit on the smallest scales at which light can be controlled. A nanoscale silver array not only circumvents the barrier, but steers different-coloured light to different places.

Is it possible to control the concentration of light energy on the nanometre scale? This 'nanofocusing challenge' has implications for a whole host of areas, including small-scale sensing and lasing techniques, high-resolution optical imaging, super-compact photonic circuitry and ultra-sensitive biochemical analysis. It is no surprise, therefore, that it lies at the heart of a lively body of research.

Sharp points on metallic particles and tiny gaps between the particles have already been explored extensively in the quest to generate optical hotspots. In taking on the nanofocusing challenge, de Waele *et al.*¹ choose another direction. Writing in *Nano Letters*, they show that a chain of tiny silver particles can be used to concentrate light at either of its ends, rather than an antenna concentrates radio waves. Not only that, but the position at which the light is located is determined by the light's colour, with a sharp switch in position occurring at a precise wavelength. The achievement is a notable step towards producing devices that can switch and steer light at the nanoscale.

But why not just make a better lens? Unfortunately, conventional devices for steering light, such as lenses or mirrors, no matter how perfectly made, are limited by diffraction. This means that any focusing effect is limited to the scale of the light's wavelength — a few hundred nanometres at best. Plasmonic antennas of the sort exploited by de Waele *et al.* take advantage of the fact that the light becomes coupled to the 'plasma' of free electrons that suffuses any metal. Such antennas are generally accepted as the best way to get round the limitation imposed by diffraction and so convert light into nanoscale-localized energy. Proven concepts used for radio and microwave antennas have thus been revisited and scaled down to the smaller-wavelength domain of visible light. Optical dipole² and monopole³ antennas — analogous to the most basic radio-antenna technologies — have been fabricated and used to confine green light fields (wavelength around 500 nm) to a region of just 20 nm.

So far, so simple, it would seem. But the optics of metals has its own peculiarities and attractions. Among the peculiarities is that metals are not perfect conductors at visible

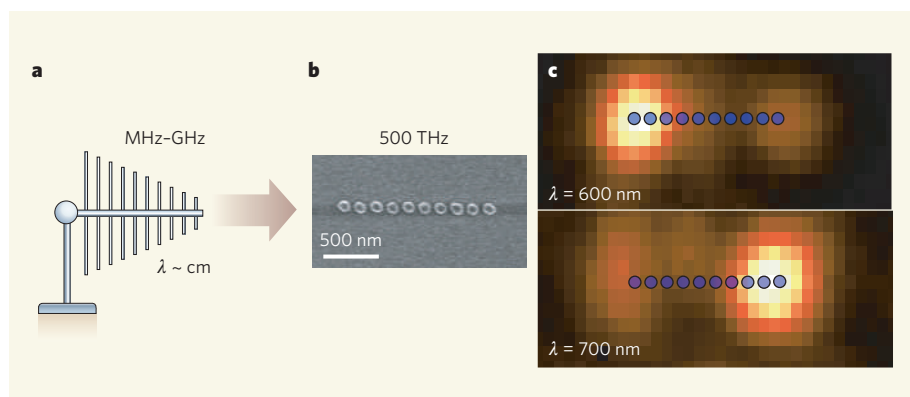


Figure 1 | Signal success. Multi-element antennas concentrate the energy from a source in a specific direction on a single element. The concept is familiar for radio waves in the megahertz–gigahertz bands (a). As de Waele *et al.*¹ show (b), it works equally well for light waves of submicrometre wavelength (frequency around 500 terahertz). In this case, the light energy is localized at the front or back of the antenna element, depending on the colour of the incident light (c), opening up a range of new applications.

wavelengths, as they are for radio waves. Instead, a metal becomes somewhat transparent: light penetrates into its surface, where oscillations of the metal's free-electron plasma take over some of the light's energy to form a resonant energy package known as a plasmon. To keep optical nanoantennas perfectly tuned, therefore, a tailored rescaling is needed that depends on the metal of choice, the shape of the antenna, the colour and polarization of the light, and so on⁴. But these complications are more than rewarded by the fact that resonant oscillation of the free-electron plasma results in strongly enhanced local fields — the desired optical hotspots — that are again determined by the metal and its geometry, and by the light's frequency. Essentially, this is an advantage, as it allows lots of creativity in designing plasmonic antennas.

With steady advances in fabrication techniques, a host of different plasmonic nanostructures has been crafted: particles, holes, wires, shells, gaps, dimers, bow-ties, arrays, to name just a few. Local optical hotspots are indeed found in most of them, mainly at edges and in gaps. The quest for active control of these hotspots has started. For example, attempts have been made to tweak the polarization of the incident light and thus

switch the position of the hotspots⁵. In a recent landmark experiment, a tiny spot of light was produced at a predefined position using a technique known as adaptive pulse control⁶, in direct analogy to the coherent control of molecular dynamics.

Chains of nanoparticles are, like multi-element radio antennas, particularly flexible structures⁷ (Fig. 1a). In these chains, both plasmon coupling and interference can be engineered to optimize photon energy transport and localization. Already, light has been guided along nanowires of silver particles much thinner than its wavelength⁸, in contravention of the diffraction limit, and its coupling to localized plasma resonances has been used to measure nanometre-scale distances⁹. Recently it has been predicted that, when a nanoparticle chain is illuminated from one direction, the plasmon interferences can sweep up the electric field from the first to the last particle along the chain, leading to a huge enhancement in the field at the chain's far extremity¹⁰.

This is the effect now exploited by de Waele and colleagues¹. In their experiment, a chain of ten silver particles, 110 nm in lateral diameter and 50 nm tall, is embedded in a homogeneous glass–oil environment and illuminated from one end (Fig. 1b). Light scattered by the chain

is viewed by a powerful confocal microscope. Remarkably, as the colour of the incident light is changed from green to red (increasing its wavelength), the area of the chain where the light is confined, as shown by the maximum amount of scattered light, flips from the front end to the back end, in a sharp transition at around 675 nm (Fig. 1c).

This extreme sensitivity of the light confinement to the incident colour points to wavelength-dependent interference effects, possibly enhanced by delayed interactions of the light with more distant parts of the nanoparticle array. Indeed, nice agreement with the measured switching is obtained in a full dynamic calculation that includes such retardation effects. The authors stress that a quasistatic approach — one that takes into account interactions only in a particle's near field — produces no asymmetric light localization.

Efficient funnelling of the light to a localized spot therefore requires proper tailoring of the interferences caused by coupling between several resonant plasmonic particles. Moreover, the fact that the widely used quasistatic approach is invalid is a severe caveat concerning the reliability of many near-field calculations on plasmonic nanostructures. In this context, it should be noted that a scattering image such as that viewed by de Waele *et al.* is dominated by areas where the plasmon is converted into emitted light, and it would be interesting to compare this image with a map of the plasmon mode in its near field.

As a final point, it is interesting to return to the comparison of the nanoparticle array to the traditional dipole-array (Yagi Uda) antenna (Fig. 1a) used to pick up radio or television signals from a station in a specific direction. The spacing between the elements in such an antenna is optimized for constructive interference to enhance the directional sensitivity and concentrate the response at the final element — in full analogy to de Waele and colleagues' optical array¹. Yet there is a significant difference: the nanoparticle antenna relies on electron-plasma oscillations that can be tuned and exploited independently. Moreover, the optical antenna is resonant at wavelengths similar to those of the optical transitions of molecules and quantum dots. One can expect that this resonant coupling will soon be exploited to produce ultra-small, wavelength-sensitive directional sensors or emitters based on nanoparticle chains. In this dynamic field, stay tuned for new surprises soon. ■

Niek F. van Hulst is at ICFO – The Institute of Photonic Sciences, Mediterranean Technology Park, Avinguda del Canal Olímpic, 08860 Castelldefels (Barcelona), and ICREA – Institució Catalana de Recerca i Estudis Avançats, 08015 Barcelona, Spain.
e-mail: niek.vanhulst@icfo.es

1. Taminiau, T. H. *et al.* *Nano Lett.* **7**, 28–33 (2007).
2. Novotny, N. *Phys. Rev. Lett.* **98**, 266802 (2007).
3. Imura, K., Okamoto, H., Hossain, M. K. & Kitajima, M. *Nano Lett.* **6**, 2173–2176 (2006).
4. Aeschlimann, M. *et al.* *Nature* **446**, 301–304 (2007).
5. Krenn, J. R. *et al.* *Phys. Rev. Lett.* **82**, 2590–2593 (1999).

6. Maier, S. A. *et al.* *Nature Mater.* **2**, 229–232 (2003).
7. Sonnichsen, C., Reinhard, B. M., Liphardt, J. & Alivisatos, A. P. *Nature Biotechnol.* **23**, 741–745 (2005).
8. Ghenuche, P., Cormack, I. G., Badenes, G., Loza-Alvarez, P. & Quidant, R. *Appl. Phys. Lett.* **90**, 041109 (2007).

GENE TRANSCRIPTION

Extending the message

Patrick Cramer

During transcription, RNA polymerase catalyses the addition of nucleotides to the growing RNA chain. High-resolution structural snapshots indicate that the polymerase first identifies its substrate, and then incorporates it.

Life is chemistry — well, at least it is to the molecular biologist. But chemical details remain unclear even for some of the fundamental cellular processes such as gene transcription. In two papers published on pages 157 and 163 of this issue, however, Vassylyev and colleagues^{1,2} provide detailed structural insights into the process of transcription and suggest a two-step mechanism for adding building blocks to a growing RNA chain.

In all living cells, gene transcription is the first step in the decoding of genetic information. During transcription, the enzyme RNA polymerase (RNAP) moves along a DNA template and synthesizes a complementary chain of ribonucleotides — the messenger RNA. Extension of the RNA chain begins with the binding of a nucleoside triphosphate (NTP) substrate to the transcription elongation complex, which consists of RNAP, DNA and RNA. Catalytic addition of the nucleotide to the growing end of the RNA chain then releases a pyrophosphate ion. Finally, translocation of DNA and RNA over the RNAP surface frees the substrate-binding site for the next NTP.

In previous attempts to understand the mechanism of transcription, structural information on the elongation complexes containing eukaryotic RNAPII or bacterial RNAP was obtained^{3–7}. These studies showed that DNA enters a cleft-like channel in RNAP. A short hybrid duplex then forms between the DNA template strand and the RNA product above the active site of RNAP at the floor of the cleft. These studies had also indicated possible mechanisms both for the separation of the two DNA strands close to the active site and for RNA–DNA separation at the end of the hybrid.

In some elongation-complex structures, NTP-binding sites were also observed within RNAPII (refs 4–6). The NTP substrate was found to be trapped in the insertion site of RNAP^{4,6} — which is apparently occupied during the catalytic extension of the RNA chain — but also in an overlapping position that slightly differed, indicating the existence of a catalytically inactive NTP-bound state⁵.

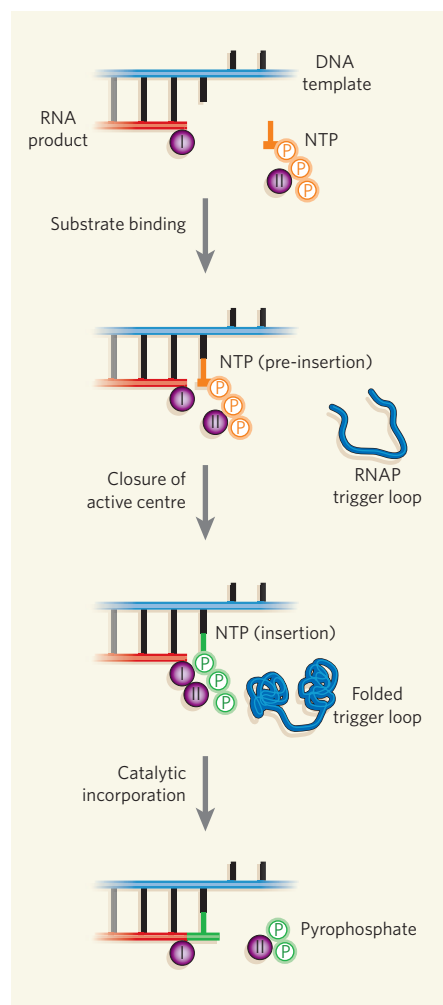


Figure 1 | Two-step mechanism of RNA chain elongation during transcription. Vassylyev *et al.*^{1,2} found that, during transcription from a DNA template, elongation of an RNA sequence involves the binding of a nucleoside triphosphate (NTP) to the RNA polymerase (RNAP) enzyme in an inactive 'pre-insertion' state; this is characterized by partial folding of the RNAP trigger loop. Complete folding of the trigger loop brings the two metal ions, I and II, of RNAP close together, delivering the NTP to the insertion site, where catalysis leads to nucleotide incorporation and the release of pyrophosphate.

1. de Waele, R., Koenderink, A. F. & Polman, A. *Nano Lett.* **7**, doi:10.1021/nl070807q (2007).
2. Mühlischlegel, P., Eisler, H.-J., Martin, O. J. F., Hecht, B. & Pohl, D. W. *Science* **308**, 1607–1609 (2005).

However, the earlier studies suffered from relatively low resolutions, ranging from 3.5 to 4.5 Å.

In one of their two papers¹, Vassylyev *et al.* now present the crystal structure of an elongation complex containing bacterial RNAP at 2.5 Å — the highest-resolution structure of an elongation complex for a cellular RNAP, and the first crystal structure of an elongation complex containing bacterial RNAP. They find that the arrangement of DNA and RNA in the complex is very similar to that of RNAPII-containing elongation complexes in higher organisms; this reflects a conserved mechanism of transcriptional elongation throughout evolution.

The authors also imaged an RNA single strand within an exit tunnel that leads to the enzyme surface. The exit tunnel could widen and accommodate an RNA duplex, which forms a hairpin structure during pausing and termination of transcription. The elongation complex structure further shows that DNA strands are separated just before the active site, allowing only one NTP to bind at a time.

The new structure of the bacterial elongation complex allowed Vassylyev and colleagues to carry out detailed analysis of the mechanism of nucleotide incorporation into the growing RNA chain in a separate study². They determined two crystal structures of the bacterial elongation complex bound to a non-reactive NTP analogue at a resolution of 3.0 Å — one in the absence and one in the presence of the antibiotic streptolydigin.

In the absence of streptolydigin, NTP binds to two catalytic metal ions at the insertion site, in agreement both with structural and functional information for RNAPII (refs 4,8,9) and with a general two-metal-ion mechanism of action for polymerases¹⁰. Metal ion I (or A) is persistently bound to RNAP, whereas metal ion II (or B) binds the triphosphate moiety of NTP, and is recruited ad hoc. When NTP binds to the insertion site, a mobile part of the active centre called the trigger loop folds completely^{2,6}.

By contrast, in the presence of streptolydigin, the trigger loop folds only partially, and the NTP binds to RNAP in a catalytically inactive 'pre-insertion' state. In this state, the base and sugar groups of NTP are essentially bound as they normally bind during insertion, but its triphosphate group and metal ion II are too far from metal I to allow catalysis. Thus, the antibiotic apparently inhibits transcription by preventing complete folding of the trigger loop, thereby impairing NTP delivery to the insertion site.

The two overlapping but chemically distinct NTP-binding sites in the RNAP active centre support a two-step mechanism of nucleotide incorporation into the growing RNA chain (Fig. 1). The NTP would first bind to an open-conformation active centre in an inactive pre-insertion state. Complete folding of the trigger loop then leads to the closure of the

active centre, delivery of the NTP to the insertion site, and the formation of all contacts required for catalysis. Biochemical data strongly support the existence of the pre-insertion state, as mutations that interfere with folding of the trigger loop hardly affect NTP binding, but they slow catalysis, apparently because NTP delivery to the insertion site is impaired².

The two-step mechanism helps explain transcription fidelity. Selection of the correct NTP would begin by sampling of substrates in the pre-insertion state, when Watson–Crick base pairing between the correct NTP and the DNA template is established, and RNAP contacts discriminate between an NTP and a deoxy-NTP, to prevent DNA synthesis. The appropriate NTP is then delivered to the insertion site, where its correct base pairing with the template is double-checked by a tight fit in the complementary closed active centre. Subsequent catalysis leads to RNA extension and pyrophosphate formation. It could be the release of pyrophosphate that destabilizes the closed conformation of the active centre¹¹, leading to the unfolding of the trigger loop, and enabling translocation of the nucleic acids.

How exactly translocation occurs, however, remains an open question. Another outstanding puzzle is how various external factors

that regulate transcription influence RNA synthesis. By mutating specific amino-acid residues in RNAP, researchers should now be able to further define the transcription mechanism and the functions of the trigger loop and other components of this enzyme. The results of Vassylyev and colleagues^{1,2} provide new clues on how to design such mutations and how to address the remaining questions.

Patrick Cramer is at the Gene Center Munich, Department of Chemistry and Biochemistry, Ludwig-Maximilians-Universität München, Feodor-Lynen-Straße 25, 81377 München, Germany.
e-mail: cramer@lmb.uni-muenchen.de

1. Vassylyev, D. G., Vassylyeva, M. N., Perederina, A., Tahirov, T. H. & Artsimovitch, I. *Nature* **448**, 157–162 (2007).
2. Vassylyev, D. G. *et al.* *Nature* **448**, 163–168 (2007).
3. Gnatt, A. L., Cramer, P., Fu, J., Bushnell, D. A. & Kornberg, R. D. *Science* **292**, 1876–1882 (2001).
4. Westover, K. D., Bushnell, D. A. & Kornberg, R. D. *Cell* **119**, 481–489 (2004).
5. Kettenberger, H., Armache, K.-J. & Cramer, P. *Mol. Cell* **16**, 955–965 (2004).
6. Wang, D., Bushnell, D. A., Westover, K. D., Kaplan, C. D. & Kornberg, R. D. *Cell* **127**, 941–954 (2006).
7. Korzhova, N. *et al.* *Science* **289**, 619–625 (2000).
8. Sosunov, V. *et al.* *EMBO J.* **22**, 2234–2244 (2003).
9. Cramer, P., Bushnell, D. A. & Kornberg, R. D. *Science* **292**, 1863–1876 (2001).
10. Steitz, T. A. *Nature* **391**, 231–232 (1998).
11. Yin, Y. W. & Steitz, T. A. *Cell* **116**, 393–404 (2004).

EXTRASOLAR PLANETS

Water on distant worlds

Heather A. Knutson

Is the presence of water a feature common to all gas-giant planets? The first convincing detection of water vapour in the atmosphere of such a planet from outside our Solar System indicates that the answer is yes.

Gas-giant planets — Jupiter, Saturn, Uranus and Neptune in our own Solar System — generally form at large distances from their parent star. Here, radiation is less intense, and so water and other low-mass elements and compounds, such as methane, can be more easily accreted in the form of ices onto the newly formed protoplanet. This process explains why the Solar System's gas giants contain significantly larger quantities of water than does the relatively rocky planet we call home. The rule is expected to hold for gas giants outside our Solar System, which we presume formed through similar processes. Until recently, however, the only planets we were able to study in the necessary detail were those in our own backyard.

On page 169 of this issue, Tinetti and colleagues¹ provide the first convincing evidence for water vapour in the atmosphere of an extrasolar planet. The hot gas-giant planet, or 'hot Jupiter', that they studied, known as HD 189733b, is too far away to be imaged directly. Instead, the authors took advantage of the unusual geometry of the system, which is

oriented such that the planet eclipses its parent star once every orbit. This transiting geometry of HD 189733b was exploited most recently to produce the first-ever infrared 'map' of the temperatures in an extrasolar planet's atmosphere^{2,3}. Because, unlike the giants of our Solar System, HD 189733b orbits extremely close to its parent star — at less than 3% of the Earth–Sun distance — these temperatures range from a toasty 1,200 kelvin on the dayside of the planet to a relatively balmy 970 kelvin on the nightside.

Tinetti *et al.* used NASA's powerful Spitzer Space Telescope to look for the signal of water absorption in starlight transmitted through the edges of the planet's atmosphere during transit. This method was applied successfully a few years ago to detect sodium in the atmosphere of a hot Jupiter⁴. The authors find that the effective light-blocking area of HD 189733b is slightly larger when measured at a wavelength of 5.8 µm than at 3.6 µm. This difference, they conclude, is the effect of water vapour in the planet's atmosphere, which will absorb more

light at the longer wavelength and transmit more at the shorter wavelength.

The detection comes as a relief for the theorists who had predicted^{5–8} that water vapour should be a significant component of the atmospheres of hot Jupiters. But it contradicts previous studies of the same planet⁹ and of another hot Jupiter¹⁰, both of which found no evidence for water. The earlier study of HD 189733b involved looking for evidence of water absorption in the spectrum of light emitted by the hot dayside of the planet⁹ (HD 189733b always presents the same face to its parent star).

So how can these findings¹ and the seemingly contradictory conclusions of the earlier work⁹ be reconciled? A previous theoretical study⁷ might hold the key: owing to the intense radiation from the star, the temperature of HD 189733b's atmosphere might remain constant over a range of altitudes (unlike Earth's atmosphere, which gets colder the higher you go). Absorption lines in a spectrum are created when radiation from a hot source (the interior of the planet in the earlier study) travels through a layer of cooler gas, where molecules selectively absorb light at a few frequencies. If the interior of the planet is the same temperature as the outer layers, this absorption doesn't occur, and the spectrum of light emitted from the dayside of the planet will be featureless.

This effect would wash out the signature of water in the dayside emission spectrum observed in the first study⁹, even if a large amount of water were present. Tinetti *et al.*¹, however, measured light from the star that then passed through the outer layers of the planet's atmosphere. Because the star is always much hotter than the planet, the spectrum of this light will always have absorption lines, regardless of the temperature profile in the planet's atmosphere. The authors' measurement is also consistent with a previous claim¹¹ that water had been detected in the transmission spectrum of another hot Jupiter, HD 209458b. However, instrument effects in these data could have created the same signal, making this claim merely suggestive. The new data, by contrast, provide solid evidence.

But the data do contain a big surprise: the strength of the observed signal is almost four times larger than predicted by current models of such planets' atmospheres^{1,12}. Tinetti and colleagues argue that, rather than requiring a significantly higher abundance of water in the planet's atmosphere, this strong absorption signal can be explained by including the full range of predicted absorption lines for water molecules at the high temperatures found in HD 189733b's atmosphere². Previous models have used only those absorption lines measured in laboratory experiments, which were performed at significantly lower temperatures and therefore did not include some of the effects, such as interactions between water molecules, that are predicted to occur at these high temperatures.

Although this detection is a significant step forward, there is still much that we do not know about HD 189733b and similar hot gas-giant planets. The discovery of gas-giant planets so close to their parent stars was a great surprise to the astronomical community. Until then, it had been assumed that most planetary systems would look similar to our own, with smaller, rocky planets closer in and giant gaseous planets only farther out.

We still do not entirely understand how these massive planets migrate into such close orbits around their parent stars, or the effect that their extreme environment has on their structure. The discovery of water highlights one area of similarity between these unusual planets and the gas giants of our own Solar System. But the surprising strength of the detected signal is yet another demonstration that what

we find is never quite what our models might lead us to predict.

Heather A. Knutson is in the Department of Astronomy, Harvard University, 60 Garden Street, MS 10, Cambridge, Massachusetts 02138, USA.

e-mail: hknutson@cfa.harvard.edu

1. Tinetti, G. *et al.* *Nature* **448**, 169–171 (2007).
2. Knutson, H. A. *et al.* *Nature* **447**, 183–186 (2007).
3. Burrows, A. *Nature* **447**, 155–156 (2007).
4. Charbonneau, D. *et al.* *Astrophys. J.* **568**, 377–384 (2002).
5. Seager, S. *et al.* *Astrophys. J.* **632**, 1122–1131 (2005).
6. Barman, T. S., Hauschildt, P. H. & Allard, F. *Astrophys. J.* **632**, 1132–1139 (2005).
7. Fortney, J. J. *et al.* *Astrophys. J.* **652**, 746–757 (2006).
8. Burrows, A., Sudarsky, D. & Hubeny, I. *Astrophys. J.* **650**, 1140–1149 (2006).
9. Grillmair, C. J. *et al.* *Astrophys. J.* **658**, L115–L118 (2007).
10. Richardson, L. J. *et al.* *Nature* **445**, 892–894 (2007).
11. Barman, T. S. *Astrophys. J. Lett.* (in the press); preprint available at www.arxiv.org/astro-ph/0704.1114 (2007).
12. Tinetti, G. *et al.* *Astrophys. J.* **654**, L99–L102 (2006).

PLANT ECOLOGY

Family roots

Ragan M. Callaway and Bruce E. Mahall

Experiments in which related and unrelated plants were grown together reveal the ability of roots to recognize their kin. The ecological and evolutionary implications are tantalizing topics for future studies.

Knowing who are your relatives and who are not creates behavioural, ecological and evolutionary opportunities. Organisms capable of recognizing kin can adjust territories, avoid incestuous mating, decide to fight or not and, importantly, benefit evolutionarily from promoting the success of brothers, sisters and cousins — individuals that share their genes. Kin recognition is widespread in animals that recognize relatives by sight, hearing, smell and even taste. Dudley and File¹, writing in *Biology Letters*, now demonstrate that plants can also recognize their family members — through their roots.

Plants do not see, hear, smell or taste. But they communicate with each other in other ways, including chemical signalling among roots²; unknown mechanisms for pollen–stigma recognition; chemical signals between parasitic plants and their hosts³; volatile molecules emitted as warning signals by leaves damaged by attack⁴; gases in smoke from burning plants⁵; and neighbour-altered light-wavelength ratios⁶. Some of these communication techniques function to discriminate between self and non-self⁷, and between members of the same population and those of a distant population⁸. However, unequivocally documenting self versus non-self recognition has proved difficult⁹, and clear demonstrations of recognition of relatedness among plants are still rarer. If plants, like animals, can distinguish related non-self from unrelated non-self “and adjust their behaviour on the basis of kinship”¹⁰, thereby enhancing the prosperity of

close relatives, their genes might preferentially be passed on to future generations in a process known as kin selection.

But to do this, kin must first be distinguished from strangers. A plant called the Great Lakes sea rocket (*Cakile edentula*; Fig. 1, overleaf) seems to possess this ability. Dudley and File¹ found that, when individuals from the same mother were planted together in a pot, the total root mass produced was less than when individuals from different mothers were planted together. It is not clear whether kin-recognition siblings reduced their own root growth, or whether unrelated individuals aggressively increased root growth to seize resources in a ‘tragedy of the commons’ scenario¹¹. Either way, roots of the sea rocket responded differently to kin than to strangers. A possible result of lower root mass in the presence of kin is less root overlap, which could mean less competition for resources among kin than among strangers.

Root behaviour⁸, overlap and competition were not directly measured by Dudley and File, however, leaving the ecological and evolutionary consequences in question. As Hans De Kroon remarked to one of us, for kin recognition to be interpreted as kin selection, one must demonstrate that recognizing kin results in increased progeny among the group of relatives — that is, in increased fitness. The plants in Dudley and File's experiment showed no such reproductive increase, but this may have been because they were grown in pots. Pots can be good for detecting root recognition but



50 YEARS AGO

"Dr. Josephine Macalister Brew, C.B.E." — It is hard to believe that Dr. Macalister Brew is dead; for if any one adjective could have described her it would have been the hackneyed word 'vital'... [but it] was not surprising that she was worked to death. Government departments, charitable trusts, bodies as varied as the Marriage Guidance Council and the Educational Drama Association all made demands on her strength; and to none of them did she give half-measure.... there must be scores [of boys and girls] who remember this odd little figure who knew what they were thinking before they did and, more, could put it into intelligible words.

From *Nature* 13 July 1957.

100 YEARS AGO

The problem of determining the motion of the sun amongst the stars has undergone a great change in consequence of Prof. J. C. Kapteyn's investigations... These researches indicated that the stars surrounding us do not form a simple system, but a dual one. From a discussion of the motions of the stars of Bradley's catalogue, Prof. Kapteyn demonstrated the existence of two great streams of stars passing through one another, and found the directions of motions of these streams relative to the sun and to one another. The Bradley stars, numbering about 2600, are mainly stars visible to the naked eye; they cover nearly three-quarters of the celestial sphere, and throughout the whole of this area Prof. Kapteyn found the two streams prevailing, and it seemed probable that all the stars he examined belonged to one or other of the two streams... In conclusion, whilst Prof. Kapteyn's theory accounts in a simple manner for the very anomalous and unsymmetrical way in which the directions of motion of the stars are distributed, it is still awaiting the verdict of the spectroscopic determinations of line-of-sight velocities.

A. S. Eddington

From *Nature* 11 July 1907.



Figure 1 | The Great Lakes (or American) sea rocket, *Cakile edentula*. Dudley and File's experiments¹ on root growth show that kin recognition occurs in this species. Whether kin recognition translates into kin selection in sea rocket remains an open question.

they are lousy for assessing a plant's full growth potential. Avoidance of root overlap, as a consequence of recognition, requires space, and space is severely restricted in pots.

In a natural field setting, however, a plant detecting neighbouring kin at low densities could decrease root growth in the vicinity of neighbours and increase root growth away from them, thereby decreasing root overlap and competition for resources and increasing its performance. In support of this possibility, other research on sea rocket has shown that, in the field, groups of siblings have higher reproduction rates than groups of strangers¹². By demonstrating kin recognition, Dudley and File have taken the critical first step; measuring the fitness consequences should be comparatively routine.

Kin recognition is sometimes linked to altruism, but reduction in root growth and overlap among sibling neighbours may not be purely altruistic. By detecting family members and sharing space with them, sea rockets may simply be ensuring that direct competition for resources does not suppress all members of the group. Competition among densely spaced individuals could result in a limiting resource being spread so evenly among relatives that no individual acquires enough to reproduce. This would be a disaster for an annual species such as sea rocket. On the other hand, kin recognition accompanied by inhibition⁸ of neighbouring roots may constitute a formidable interference mechanism that would allow plants to form large enough territories¹³ for the successful growth and reproduction of some family members at the expense of others. Kin recognition, therefore, may not directly benefit all members in a family. But it may increase the odds that at least a few members will successfully pass on the family genes to the next generation.

Without brains, how do plants recognize

their relatives? No one knows. Possibilities include communication through chemical exudates, released volatile molecules, electrical signals, and enzymes functioning at cell surfaces¹⁴. Research on a desert shrub, *Ambrosia dumosa*, suggests that root interactions may involve at least two levels of recognition and interplay among physiological and genetic components. At the self versus non-self level, roots on the same individual shrub did not inhibit each other upon contact, whereas roots from genetically identical but physically separated individuals did¹³. At the population level, roots on different plants from the same population inhibited each other, whereas those from different populations did not⁸. In addition, recognition and response probably constitute two different mechanisms, because studies on other species have found, in contrast, that self-recognition can lead to inhibition¹⁵.

Clearly, research on root behaviour is just beginning. But if neighbour identity commonly dictates root interactions, major overhauls of theories that assume that direct resource competition determines plant community organization will be necessary. And if kin recognition among roots can be unequivocally linked to evolutionary consequences, we will have to expand the pool of mechanisms known to drive plant evolution. ■

Ragan M. Callaway is in the Division of Biological Sciences, University of Montana, Missoula, Montana 59812, USA. Bruce E. Mahall is in the Department of Ecology, Evolution, and Marine Biology, University of California, Santa Barbara, California 93106, USA.

e-mails: ray.callaway@mso.umt.edu; mahall@lifesci.ucsb.edu

1. Dudley, S. A. & File, A. L. *Biol. Lett.* doi:10.1098/rsbl.2007.0232 (2007).

2. Mahall, B. E. & Callaway, R. M. *Ecology* **73**, 2145–2151 (1992).

3. Yoder, J. I. *Planta* **202**, 407–413 (1997).
4. Karban, R., Baldwin, I. T., Baxter, K. J., Laue, G. & Felton, G. W. *Oecologia* **125**, 66–71 (2000).
5. Keeley, J. E. & Fotheringham, C. J. *Science* **276**, 1248–1251 (1997).
6. Davis, M. H. & Simmons, S. R. *Plant Cell Environ.* **17**, 829–836 (1994).
7. Falik, O., Reides, P., Gersani, M. & Novopiansky, A. *J. Ecol.* **91**, 525–531 (2003).
8. Mahall, B. E. & Callaway, R. M. *Am. J. Bot.* **83**, 83–98 (1996).
9. Hess, L. & De Kroon, H. *J. Ecol.* **95**, 241–251 (2007).
10. Hamilton, W. D. J. *Theor. Biol.* **7**, 1–16 (1964).
11. Gersani, M., Brown, J. S., O'Brien, E. E., Maina, G. M. & Abramsky, Z. *J. Ecol.* **89**, 660–669 (2001).
12. Donohue, K. *Am. Nat.* **162**, 77–92 (2003).
13. Schenck, J., Mahall, B. E. & Callaway, R. M. *Adv. Ecol.* **28**, 145–180 (1999).
14. Hiscock, S. J., Bown, D., Gurr, S. J. & Dickinson, H. G. *Sex. Plant Reprod.* **15**, 65–74 (2002).
15. Holzapfel, C. & Alpert, P. *Oecologia* **134**, 72–77 (2003).

NEUROPHYSIOLOGY

Channelling cold reception

Bernd Nilius and Thomas Voets

Perception of cold and hot is one of life's essentials. Three research teams find that, when a temperature-sensing receptor is deleted in mice, the animals lose their response to a range of cold temperatures.

In his description of the five senses, Aristotle described *visus* (sight) as the most supreme sense, yielding the highest pleasure, and *contactus* (touch and sensing temperature) as the most rudimentary sense, required for sheer survival¹. Indeed, to maintain a healthy core body temperature of 37 °C, humans — like other animals that can retain a relatively constant internal body temperature — need to be able to 'feel' the ambient temperature and show a suitable physiological or behavioural response to drastic fluctuations in it.

Ambient temperatures are sensed by cells of the peripheral nervous system, which convey thermal information from the skin and peripheral tissues to the brain (for reviews see refs 2,3). Three papers^{4–6}, including one by Bautista *et al.* in this issue, now report the consequences of deleting the gene encoding a peripheral cold sensor called TRPM8. These researchers find that mice that do not have the TRPM8 cation channel — a member of the transient receptor

potential (TRP) family — have severe deficiencies in sensing cold and in cold-induced behaviour.

It is not surprising that the deletion of the *TRPM8* gene leads to reduced cold sensitivity. Previous studies had shown that TRPM8 is a temperature-sensitive TRP channel that is activated by moderate cooling and by 'cool' substances such as menthol, eucalyptol and icilin³. It is expressed in the free nerve endings of a subset of small-diameter sensory neurons⁷; the nerve fibres of these neurons, which are not covered by the myelin sheath, carry sensory information from the skin to the brain (Fig. 1a). Consequently, TRPM8 had been proposed as the source of non-painful and painful reactions to cold³ and as the molecular mediator of cold-induced pain relief⁸.

The results of studies^{4–6} on TRPM8-deficient mice mainly endorse these earlier views. At a cellular level, all three studies^{4–6} showed that sensory neurons of these mice

show a drastically blunted response to cold stimuli — for example, a drop in temperature from around 30 °C to below 20 °C — and to menthol. Behavioural studies^{4–6} illustrate the consequences of such severe deficits in cold sensation. When given the freedom of choice, mice with their *TRPM8* gene intact prefer to reside in a warm (around 30 °C) rather than a cool (20 °C or lower) zone. By contrast, those without TRPM8 do not discriminate against cool temperatures, cheerfully walking into the cold. Moreover, these mice no longer exhibit the typical 'wet-dog-shake' response to injections of icilin, and show a reduced sensitivity to extreme and painful cold stimuli. Finally, Colburn *et al.*⁴ find that TRPM8 might participate in hypersensitivity to cold, which is observed after inflammation or nerve injury⁹.

So, do these studies fully elucidate the mechanism of cold sensing? Not really. First, all three papers^{4–6} equivocally report the existence of a fraction of neurons in the TRPM8-deficient mice that still respond to cold. These neurons have a low temperature threshold for activation by cold (12 °C compared with around 22 °C for TRPM8-containing neurons)⁶, and so they might be important for noxious cold sensing. The debate about whether the remaining cold sensitivity is mediated by another TRP channel, TRPA1 (ref. 3), is ongoing, and analysis of mice lacking both TRPM8 and TRPA1 should eventually settle this. Furthermore, another recent study¹⁰ reports that a non-TRP channel — the voltage-dependent sodium channel Na_v1.8 — is a candidate sensor of very low temperatures.

These observations on TRPM8-deficient mice^{4–6} also prompt a look back at neurophysiological results on the fundamental basis of temperature sensation published in the early 1950s. Hensel and Zotterman¹¹ reported the temperature dependence of cutaneous C and

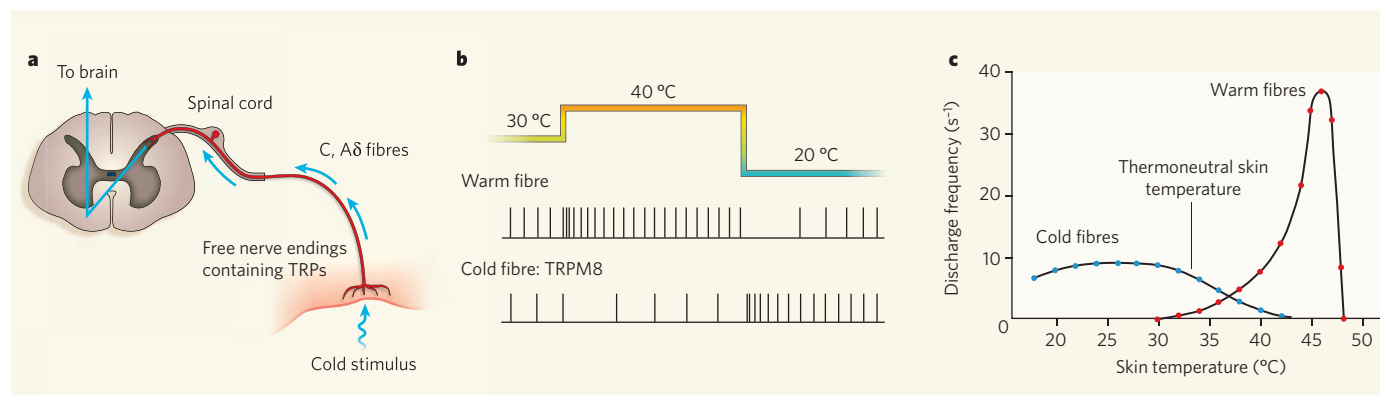


Figure 1 | Sensing cold. Three studies^{4–6} implicate the TRPM8 cation channel in cold perception. **a**, Stimulation of TRPM8 in free nerve endings of C and Aδ fibres leads to the repetitive discharge of action potentials. These signals propagate into the dorsal root of the spinal cord, and stimulate neurons that transmit thermal information to the brain. **b**, Stepwise changes in temperature evoke mirror responses in warm and cold fibres. Warm fibres expressing heat-activated, temperature-sensitive TRPs respond to sudden heating with a transient

increase in the discharge frequency of action potentials (on-response), whereas sudden cooling leads to a transient decrease in frequency (off-response). Cold fibres expressing the TRPM8 channel show an opposite reaction — an off-response to heating and an on-response to cooling. **c**, Steady-state discharge rates of cold and warm fibres. Note that the same steady-state discharge rate can occur at two different temperatures in a single fibre, raising the question of how the sensation of different temperatures is decoded.

TOXICOLOGY

Danger in the diet

Visit websites selling homeopathic remedies, and sooner or later you will find the virtues of extracts of *Aristolochia clematitis* extolled. The dark side of this plant's biochemical products is examined by Arthur Grollman, Bojan Jelaković and colleagues in *Proceedings of the National Academy of Sciences* (A. P. Grollman *et al.* *Proc. Natl Acad. Sci. USA* doi:10.1073/pnas.0701248104). They provide a strong case that the aristolochic acid produced by *A. clematitis* is the cause of endemic (Balkan) nephropathy.

This disease results in kidney failure and is associated with cancer of the upper urinary tract. Its name stems from its incidence in farming communities along tributaries of the Danube. Here, as Grollman *et al.* confirmed in their early work done in Croatia, the plant grows in wheat fields and its seeds become mixed with wheat grains

during harvesting — and so can contaminate the flour that is subsequently baked into bread.

The authors' investigations were prompted by similarities between endemic nephropathy and a condition called aristolochic acid nephropathy, which was identified in a group of women in Belgium and attributed to their use of herbal products as part of a slimming regime. Aristolochic acid reacts with DNA, and forms tell-tale biomarkers that can be used as indicators of exposure to the substance.

These biomarkers indeed turned out to be present in Croatian patients with endemic nephropathy, and malignancies of the upper urinary tract, who had long inhabited villages likely to be subject to dietary contamination. They were not seen in patients with other types of kidney disease.

Grollman *et al.* then went further,

delving into the mutational background of the cancers. They focused on the tumour-suppressor protein p53, which helps protect the genome against damage. Here they identified a large proportion of switches of the adenine-thymine nucleotide coupling in DNA to thymine-adenine. This mutational 'fingerprint' is also seen in cultured cells and in rodents treated with aristolochic acid. All in all, the authors' detective work provides enough evidence to put *A. clematitis* in the dock, if not to allow an outright conviction.

Other causes of endemic nephropathy have been considered over the years, including one that echoes the notorious case of ergotism. This disorder was eventually attributed to the action of toxins produced by fungal infection of cereals. A fungal toxin, ochratoxin A, has likewise been proposed as the agent behind endemic nephropathy, but the authors find the case for that



K. STÜBER ONLINE LIB.

to be weak. However, they estimate that only about 1 in 20 people exposed to high levels of aristolochic acid develop overt disease. They conclude, then, that there must be a large genetic component to susceptibility, and investigations of that aspect will shortly be under way.

Tim Lincoln

Aδ fibres, which are part of specific thermo-sensor neurons present in nearly all vertebrates, but are especially important in mammals (Fig. 1a). These classic studies showed that neurons containing cold receptors (C and Aδ fibres) or warm receptors (mainly C fibres) exhibit a static discharge frequency of action potentials. In neurons containing warm receptors, the discharge frequency increases steeply when the temperature rises from 30 °C to 43 °C and then falls off at higher temperatures; in those with cold receptors, the discharge frequency rises as temperature drops from 40 °C to around 25 °C, and then decreases to a stationary frequency (Fig. 1c).

Based on the input from both cold and warm fibres, the central nervous system somehow identifies temperatures below the thermo-neutral skin temperature of about 33 °C as cold, and temperatures above this as warm. An outstanding question is how the static discharge pattern relates to the activity of TRPM8 and other temperature-sensitive TRPs. Bautista and colleagues⁶ provide some insights from recordings of cutaneous C and Aδ fibres.

Usually, gradual cooling of cold-sensitive C fibres from 35 °C to 2 °C activates a burst of action potentials, with the fibres eventually adopting a residual firing rate. In mice without TRPM8, the activation phase is absent, but the residual firing is preserved. This indicates that cold-induced activation and subsequent desensitization of TRPM8 underlie the transient discharge pattern that occurs on cooling.

The instantaneous responses of warm and cold receptors to temperature are mirror images of one another. Warm receptors exhibit an on-response (increase in discharge frequency) on heating, and an off-response on cooling; the opposite is true for cold receptors (Fig. 1b). This implies that cooling evokes a dual message to the brain: an increased activity of cold-sensitive fibres and a decreased activity of warm-sensitive fibres. It might also explain the classical psychophysiological observations in Ernst Weber's 'three-bowl experiment', also called Weber's illusion¹²: coming from a bowl with cold water, water at neutral temperature feels warm; coming from a bowl with warm water, the same neutral temperature is perceived as cold.

Through single nerve-fibre recordings, Bautista *et al.*⁶ found that part of the on-response of cold receptors to a cold stimulus is due to TRPM8 activation. However, whether the off-response of these TRPM8-expressing cold fibres contributes to warm perception, or whether the closing of heat-activated thermotransducers contributes to a cold response, remains unclear.

A further crucial difference between cold and warm sensations is illustrated by another warm receptor — TRPV3. This receptor is expressed in the keratinocyte cells of the skin, which pass the signal to the sensory neurons through an unidentified messenger system¹³. Can such an indirect mechanism of nerve-fibre activation in

response to cold also be relevant to TRPM8?

Aristotle appreciated the basic importance of thermosensation for survival. It is curious, however, that these studies^{4–6} did not look into the consequences of loss of cold sensitivity on thermoregulation and core temperature. A lower core temperature may even increase the lifespan¹⁴, so the cold-indifferent TRPM8-deficient mice might live longer than their cold-fearing normal mates. Nonetheless, these studies re-ignite the excitement about TRPs once again.

Bernd Nilius and Thomas Voets are in the Department of Molecular Cell Biology, Division of Physiology, Campus Gasthuisberg, KU Leuven, Herestraat 49, B-3000 Leuven, Belgium. e-mail: bernd.nilius@med.kuleuven.be

- Bernstein, J. *Die fünf Sinne des Menschen* (Brockhaus, Leipzig, 1875).
- Voets, T., Talavera, K., Owsianik, G. & Nilius, B. *Nature Chem. Biol.* **1**, 85–92 (2005).
- Dhaka, A., Viswanath, V. & Patapoutian, A. *Annu. Rev. Neurosci.* **29**, 135–161 (2006).
- Colburn, R. W. *et al.* *Neuron* **54**, 379–386 (2007).
- Dhaka, A. *et al.* *Neuron* **54**, 371–378 (2007).
- Bautista, D. M. *et al.* *Nature* **448**, 204–208 (2007).
- McKemy, D. D., Neuhauser, W. M. & Julius, D. *Nature* **416**, 52–58 (2002).
- Proudfoot, C. J. *et al.* *Curr. Biol.* **16**, 1591–1605 (2006).
- Chung, M. K. & Caterina, M. J. *Neuron* **54**, 345–347 (2007).
- Zimmermann, K. *et al.* *Nature* **447**, 856–859 (2007).
- Hensel, H. & Zotterman, Y. J. *Neurophysiol.* **14**, 377–385 (1951).
- Schmidt, R. F. & Thews, G. *Physiologie des Menschen* 214–217 (Springer, Heidelberg, 1990).
- Lee, H. & Caterina, M. J. *Pflügers Arch. Eur. J. Physiol.* **451**, 160–167 (2005).
- Conti, B. *et al.* *Science* **314**, 825–828 (2006).

OBITUARY

Pierre-Gilles de Gennes (1932–2007)

Pioneer of soft-matter physics.

“Chacun en nous a son trésor d’images entrevues dans un instant mais jamais oubliées. Un exemple pour moi: Picasso peignant à grands traits blancs sur une vitre et filmé par Clouzot. Tout ce que j’ai essayé de dessiner laborieusement plus tard est né de ces moments là.”

“Every one of us has a treasure of images caught in glimpses but never forgotten. A personal example: Picasso painting white lines on glass using large strokes, filmed behind the glass by Clouzot. Everything that I tried painting laboriously later was born from such glimpses.” (Pierre-Gilles de Gennes, from *L’émerveillement* by Thibaut de Wurtemberg, Saint-Augustin, 1998.)

With his strikingly simple yet pioneering ideas, Pierre-Gilles de Gennes drew ‘white lines in large strokes’ that defined the physics of soft matter — liquid crystals, polymers, colloids and surfactants. He died on 18 May.

Educated at the École Normale Supérieure in Paris during 1951–55, de Gennes learned theoretical physics from the greatest masters of his time. He obtained his PhD in 1957 while at the French Atomic Energy Commission, specializing in magnetism and neutron scattering. During a stay at the University of California, Berkeley, in 1959 he studied with the solid-state physicist Charles Kittel, who taught him how to communicate ideas in physics using plain language, so avoiding the use of daunting equations.

After two years in the French navy, de Gennes founded a research group at the University of Paris, Orsay, in 1961. Working on superconductivity, he became a self-proclaimed “theoretician of the tangible” — a theoretical physicist who works closely with experimentalists. But it was in 1967 that he started on work in the area that made him famous, creating the Orsay Liquid Crystals Group. Assembling a group of specialists from the fields of optics, crystallography and magnetism, along with theoreticians, he formed a truly multidisciplinary team.

De Gennes fostered a collective research effort that is scarcely imaginable today. Papers were signed not with the names of individuals, but with the name of the group. Theoreticians would spend half their time contemplating liquid crystals under the microscope and discussing practical experiments. Researchers would often arrive in the morning to find a note from de Gennes that would launch them in yet another ground-breaking direction. Calling on his vast knowledge of physics, de Gennes drew analogies between different fields. For example, he realized that laws developed

to describe superconductivity phenomena could be used to understand phase transitions in liquid crystals.

His nomination to the prestigious Collège de France in 1971 coincided with the end of his liquid-crystal research and the beginning of his work on polymers. Over the next three decades, de Gennes gave a series of public lectures on a different subject each year — colloids, granular matter and adhesion, to name just a few. While lecturing on phase transitions, de Gennes discovered the analogy between the conformations of polymer chains and the alignment of magnetic moments at a magnetic phase transition. This so-called ‘ $n=0$ ’ theorem turned polymer physics upside-down, allowing theories of phase transitions to be applied to polymers and earning him the Nobel prize in 1991.

De Gennes was a proponent of the idea that diverse phenomena could be analysed with scaling arguments — that problems could be simplified by invoking a characteristic time and length scale. This approach helps get to the heart of a problem and guides experiments. During this period, he started to use images to explain his ideas, and found creative language to describe these images. Each theory based on scaling arguments became the subject of a drawing. For instance, he likened a polymer solution to a fishing-net and called the mesh a ‘blob’. If our focus zooms in on a blob, we see the characteristic behaviour of an isolated polymer chain. But if the perspective zooms out sufficiently, we find another limit, that of the entangled polymer chains, which are composed of strings of independent blobs. The concept of blobs allows both the statics (equilibrium behaviour) and dynamics of polymer solutions to be calculated.

Another example is de Gennes’ description of polymers adsorbed onto liquid–air or liquid–solid interfaces: he drew a fractal grid that allowed the theoretical concentration profile to be solved in a single line. Finally, his ‘reptation’ model describes how a polymer chain can snake through a tube formed by surrounding molecules. In this way, the viscoelastic properties of polymer melts can be described by a few simple equations.

In 1980, de Gennes turned his attention to the science of colloids and wetting, collaborating closely with industry. Intrigued by the process of assisted oil recovery — in which polymers or surfactants are injected into oilfields to aid the flow of oil — he plunged into the physics of wetting. By devising numerous “easy experiments”



(as he put it), de Gennes elucidated the spreading dynamics of drops on surfaces; the wetting behaviour of films; and the laws that govern aquaplaning by cars.

The final phase of de Gennes’ career began in 2002, when he joined the Curie Institute in Paris. Here he studied two topics at the forefront of biology: cellular adhesion and memory formation. He took the same approach to biology as he had to industry — faced with problems of apparently insurmountable complexity, he isolated the few essential parameters to derive minimalist theories that could be used and tested.

De Gennes pursued his research with extraordinary imagination, insatiable curiosity and an ability to grasp facts rapidly. But he also gave his time to others and helped them develop their ideas. A keen ambassador of science to the public, he generated passionate debates on subjects as diverse as “Physics and Medicine”, “Inventors” and “Primo Levi”. He inspired generations of students to pursue careers in physics and played an active role in establishing the L’Oréal–UNESCO Awards for Women in Science.

Pierre-Gilles de Gennes leaves seven children, three born to Anne-Marie de Gennes and four born to me. His scientific legacy is the enduring stamp he has left on intellectual life in France and internationally on the study of the physics of so many diverse phenomena. Thanks, Pierre-Gilles, for sharing with us your insatiable love for science.

Françoise Brochard-Wyart

Françoise Brochard-Wyart is at the Institut Curie, 26 rue d’Ulm, 75248 Paris, cedex 05, France, and at the University of Paris 6. She was a scientific collaborator of Pierre-Gilles de Gennes for 35 years.

e-mail: brochard@curie.fr

A genome-wide transgenic RNAi library for conditional gene inactivation in *Drosophila*

Georg Dietzl^{1,2}, Doris Chen¹, Frank Schnorrer², Kuan-Chung Su¹, Yulia Barinova¹, Michaela Fellner^{1,2}, Beate Gasser¹, Kaolin Kinsey^{1,2}, Silvia Oppel^{1,2}, Susanne Scheiblauer¹, Africa Couto², Vincent Marra¹, Krystyna Keleman^{1,2} & Barry J. Dickson^{1,2}

Forward genetic screens in model organisms have provided important insights into numerous aspects of development, physiology and pathology. With the availability of complete genome sequences and the introduction of RNA-mediated gene interference (RNAi), systematic reverse genetic screens are now also possible. Until now, such genome-wide RNAi screens have mostly been restricted to cultured cells and ubiquitous gene inactivation in *Caenorhabditis elegans*. This powerful approach has not yet been applied in a tissue-specific manner. Here we report the generation and validation of a genome-wide library of *Drosophila melanogaster* RNAi transgenes, enabling the conditional inactivation of gene function in specific tissues of the intact organism. Our RNAi transgenes consist of short gene fragments cloned as inverted repeats and expressed using the binary GAL4/UAS system. We generated 22,270 transgenic lines, covering 88% of the predicted protein-coding genes in the *Drosophila* genome. Molecular and phenotypic assays indicate that the majority of these transgenes are functional. Our transgenic RNAi library thus opens up the prospect of systematically analysing gene functions in any tissue and at any stage of the *Drosophila* lifespan.

Geneticists have traditionally sought to gain insight into complex biological processes through forward genetic screens. Mutations are generated at random, phenotypes of interest are scored, and the mutated gene is subsequently identified. This approach has been remarkably successful, but is limited by inherent biases in mutagenesis techniques, the large numbers of mutants that must be analysed, and the considerable effort that is still required to identify the relevant genetic lesions. Moreover, most genes have multiple functions, and a gene's function in one tissue can preclude its recovery in screens focused on functions in other tissues. This is particularly true for genes that are essential in the early development of the organism.

The inhibition of gene function by RNAi¹, coupled with the availability of annotated genome sequences, now enables systematic surveys of gene function by reverse genetics. One by one, the function of almost every predicted gene can be disrupted and the phenotypic consequences observed. Any phenotype is immediately linked to a specific DNA sequence. This method has been successfully used in genome-wide screens by applying double-stranded RNAs to *Drosophila melanogaster*² or mammalian^{3,4} cells in culture. These cell-based assay systems enable detailed studies of many basic cellular processes, but not the complex biology of whole organisms. Large-scale RNAi-based surveys of gene function *in vivo* have thus far been limited to the nematode *Caenorhabditis elegans*^{5–8} and the planarian *Schmidtea mediterranea*⁹. However, in these organisms, RNAi is systemic^{1,10}, and so gene interference cannot easily be restricted to a specific cell type.

In *Drosophila*, RNAi is cell autonomous^{11,12} and can be triggered by the expression of a long double-stranded 'hairpin' RNA from a transgene containing a gene fragment cloned as an inverted repeat^{13–16}. Using the binary GAL4/UAS expression system¹⁷, such RNAi transgenes can be used flexibly to target gene inactivation to potentially any desired cell type at any stage of the animal's lifespan. If

a genome-wide library of transgenic RNAi strains were available, it would thus be possible to conduct systematic RNAi screens targeted to specific cell types in the intact animal. Here we report the generation and validation of such a library.

A genome-wide transgenic RNAi library

We first constructed a library of UAS-driven inverted repeat (UAS-IR) transgenes by cloning short gene fragments as inverted repeats into a modified pUAST vector¹⁷, pMF3 (Fig. 1a). PCR primers were designed to amplify fragments from every predicted protein-coding gene in Release 4.3 of the *Drosophila* genome sequence. Where possible, we targeted a single coding exon common to all predicted transcripts of a given gene, and used genomic DNA as a PCR template (77.8%). For some genes it was necessary to include multiple exons, or 5' or 3' UTRs, in which case mixed-stage complementary DNA was used as a template (22.2%). We successfully cloned 15,072 UAS-IR constructs, representing 13,327 different genes (97.0% of predicted protein-coding genes; Supplementary Table 1). The size of the gene fragment in each inverted repeat varies from 109 to 415 base pairs (bp), with a mean of 323 bp (Fig. 1b).

The inverted repeat sequence for each construct was predicted using e-PCR¹⁸ with low stringency criteria. In most cases (98.4%), a single product was predicted; only 380 primer pairs were predicted to amplify multiple products. We directly sequenced both halves of the inverted repeat for all 380 ambiguous primer pairs, as well as 192 of the unambiguous pairs. This confirmed that the intended product had been cloned for all of the unambiguous pairs, and all but 20 of the ambiguous pairs. These 20 anomalous clones consisted of different PCR products cloned head-to-head, and were discarded. Amongst the remaining 552 sequence-verified RNAi constructs, 8 had a deletion in one or both halves of the inverted repeat (ranging in size from 18 to 100 bp); the remaining 544 corresponded exactly to the

¹Institute of Molecular Biotechnology of the Austrian Academy of Sciences (IMBA), Dr. Bohr-Gasse 3-5, A-1030 Vienna, Austria. ²Research Institute of Molecular Pathology (IMP) Dr. Bohr-Gasse 7, A-1030 Vienna, Austria.

predicted sequence. We thus estimate that ~98.5% of the RNAi constructs contain the complete intended inverted repeat, whereas ~1.5% contain inverted repeats that are slightly smaller than intended but still likely to trigger RNAi.

The validated *UAS-IR* constructs were then used to generate a library of transgenic RNAi strains by germline transformation¹⁹ of an isogenic *w¹¹¹⁸* host²⁰. Each insertion was verified by amplification of the appropriate-sized product from genomic DNA by PCR with a vector- and gene-specific primer pair (Fig. 1a). We mapped each verified insertion to a specific chromosome and established homozygous or balanced stocks, retaining the isogenic *w¹¹¹⁸* background throughout. In all, we obtained 22,270 transgenic RNAi lines, representing 13,251 RNAi constructs and 12,088 genes (88.0% of predicted protein-coding genes; Supplementary Table 2).

Target predictions

To assess the targeting potential of each RNAi construct, we conceptually 'diced' the predicted hairpin RNA into all possible 19-mers because RNAi-mediated degradation of a target messenger RNA generally requires a perfect match of at least 19 nucleotides²¹. We then interrogated the predicted *Drosophila* transcriptome for all perfect matches to these 19-mers, in both the sense and antisense orientations. By design, all 19-mers from a *UAS-IR* construct match its primary target gene. In general, it will also be both desirable and unavoidable to additionally target any other gene with very high sequence similarity. Accordingly, we defined an on-target gene as any gene hit by at least 80% of a construct's 19-mers. Any gene hit

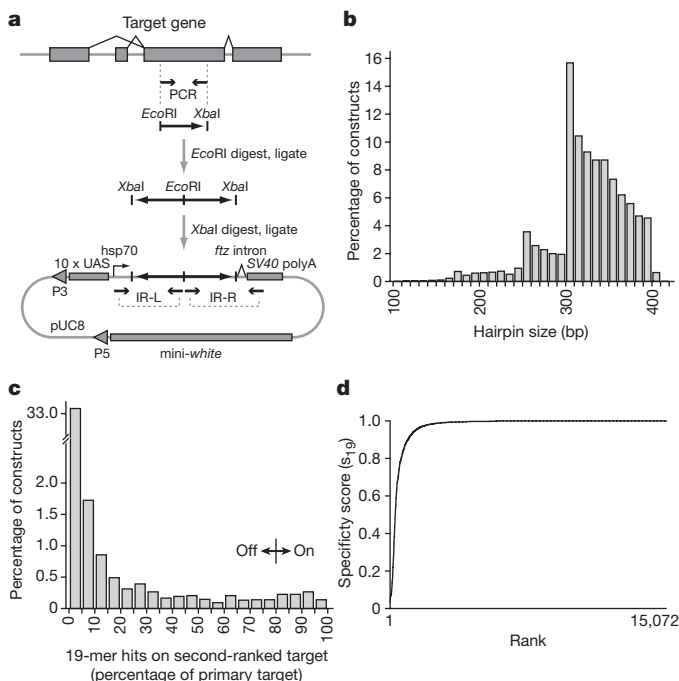


Figure 1 | A genome-wide transgenic RNAi library. **a**, Strategy used to generate *UAS-IR* constructs. Restriction sites in the original PCR primers were used to digest and ligate PCR products, followed by ligation of the inverted repeat into the pUAS vector. pUAS contains 10 GAL4-responsive *UAS* elements, the basal *hsp70* promoter, the 150 bp second intron of *ftz* and the SV40 polyadenylation signal. Most, but not all, inverted repeats were cloned in the antisense-sense orientation using *EcoRI* and *XbaI* as indicated. IR-L and IR-R indicate the primer pairs used to amplify the left or right halves, respectively, of the inverted repeat. P3 and P5 indicate P-element ends. **b**, Inverted repeat size distribution in 10 bp bins. **c**, Distribution of second-ranked targets, binned according to number of 19-mer matches (primary target = 100%). Any gene hit by less than 80% of 19-mers was considered an off-target hit. For 9,134 constructs (60.6%), the only hit was the primary target. **d**, s_{19} specificity scores for all 15,072 constructs, ranked by specificity.

by fewer 19-mers, but at least one, is considered a potential off-target gene. By these criteria, 14,612 constructs (96.9%) have a single on-target gene. Less than half of the *UAS-IR* constructs have off-target genes (5,889; 39.1%), and most of these are hit by <1% of the construct's 19-mers (Fig. 1c).

As a further measure of targeting specificity, we defined a specificity score, s_{19} , as the number of all on-target 19-mer matches divided by the total number of matches (that is $s_{19} = \Sigma \text{ on-target matches} / (\Sigma \text{ on-target matches} + \Sigma \text{ off-target matches})$). Thus, $s_{19} = 1$ for a *UAS-IR* construct with no off-target hits, and $s_{19} = 0$ for a (hypothetical) construct with only off-target hits. In our *UAS-IR* collection, $s_{19} \geq 0.80$ for 14,365 constructs (95.3%) (Fig. 1d). Our experience until now suggests that this is likely to be a fairly conservative threshold, as we have observed specific effects even with constructs having a much lower s_{19} score. Another potential source of off-target effects are CAN repeats, with double stranded RNAs containing more than 13 consecutive CANs prone to cause non-specific effects²². Only 119 (0.8%) of our RNAi constructs include more than 10 consecutive CANs (Supplementary Table 1).

Targeting in vivo

To test the efficiency of RNA knock-down, we selected a set of 64 *UAS-IR* lines (see Methods) and prepared total RNA from *Act5C-GAL4/UAS-IR* animals and *Act5C-GAL4/+* controls. RNA levels for the target gene in the knock-down and control adults were measured using quantitative PCR with reverse transcription (RT-PCR). RNA levels were reduced to less than 25% of the corresponding controls in 25 of the 64 lines (Fig. 2a and Supplementary Table 3). This is likely to underestimate the typical RNA knock-down effect, because we were restricted for this test set to lines that are viable in combination with *Act5C-GAL4*. We presume that RNA levels are generally even more strongly reduced in those lines that are lethal with *Act5C-GAL4* (~32%, see below). It is also unlikely that *Act5C-GAL4* is highly expressed in all cells at all times, and so the efficiency of knock-down in individual cells may well exceed this whole-animal estimate.

As a functional test of RNA knock-down, we selected a positive control set of 432 genes for which classical genetic methods have documented a lethal or visible mutant phenotype. For these genes, we had 658 transgenic RNAi lines in our collection (selecting only those with ≤ 5 CAN repeats and $s_{19} \geq 0.80$). We also selected 499 genes at random, for which we had 723 RNAi lines. Each of these lines was crossed in duplicate to *Act5C-GAL4* and the progeny screened for lethality or any of 159 distinct defects in adult morphology (Supplementary Table 4). For the positive control set, we observed a phenotype with at least one line for 282 genes (65.3%), and in total for 372 lines (60.1%) (Fig. 2b). Examples of morphological phenotypes are shown in Fig. 2c, along with the corresponding loss-of-function mutants as controls. For the random set, 186 genes (37.3%) covered by 225 lines (31.9%) produced a phenotypic defect (Fig. 2b).

From the positive control set, we estimate the overall false-negative rate for our library at 39.9% of lines and 34.7% of genes. These false-negative rates can almost certainly be reduced by more detailed phenotypic analyses, the use of different *GAL4* driver lines, and methods to enhance RNAi potency, as shown below. Assessing false-positive rates is much more problematic because it is difficult to reliably identify negative control genes. We could however define a negative control set of 48 genes for which classical genetics has provided strong evidence that null mutations result in either no phenotype or a specific visible phenotype. We tested the 63 lines we had for these genes with *Act5C-GAL4*, scoring them for lethality and the same set of 159 morphological defects as the positive controls and randomly selected lines. An unexpected phenotype was observed for only one line (Fig. 2b). We thus estimate that the false-positive rate is likely to be below 2% of lines. It should, however, be noted that we selected lines with ≤ 5 CAN repeats and $s_{19} \geq 0.80$ for all of these analyses. We have not systematically tested the small fraction of *UAS-IR* constructs

predicted to have a high risk of off-targeting, but anticipate that false-positive rates will be significantly higher for these constructs^{21,22}.

Tissue-specific RNAi

The availability of a diverse set of *GAL4* drivers²³ allows the transgenic RNAi library to be used to target gene inactivation to almost any desired cell type. Unlike conventional mosaic strategies using classical mutations, transgenic RNAi is not restricted to cells that are clonally related. To test the feasibility of using our library in such experiments, we selected 50 lines that were lethal in combination with the *Act5C-GAL4* driver and crossed each of them separately to *GAL4* drivers that target gene interference to the wing (*MS1096-GAL4*; an enhancer trap in the *Bx* gene), eye (*ey-GAL4*, *GMR-GAL4*), or notum (*pnr-GAL4*). Approximately 30% of the lines were also lethal with each of these drivers, most probably owing to a vital gene function in the target tissue or in other cells in which the driver is additionally expressed. However, the majority of lines were viable with the tissue-specific drivers, and 25–35% resulted in specific morphological defects (Fig. 3a, b). Different sets of lines produced phenotypes with different *GAL4* drivers, such that 80% of the lines produced a lethal or visible phenotype with at least one driver. These data illustrate the power of transgenic RNAi to uncover tissue-specific functions of essential genes.

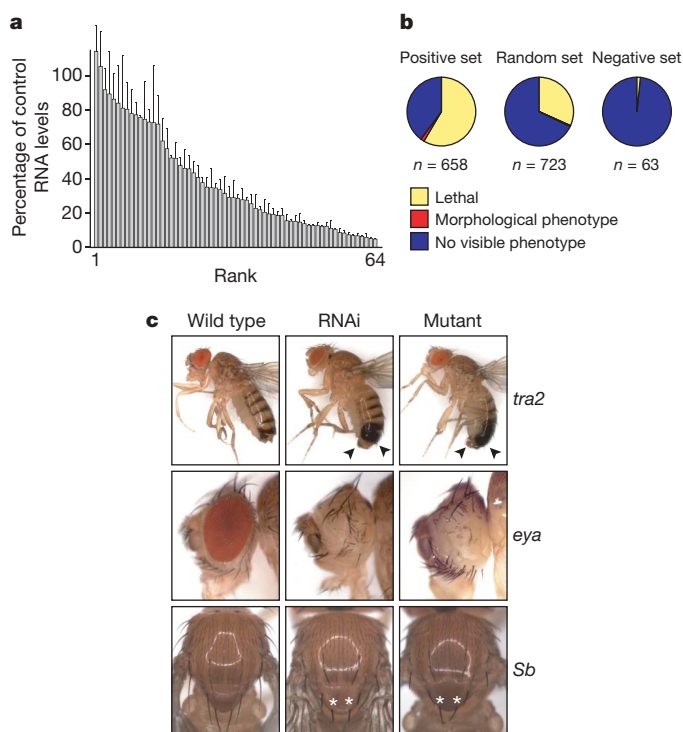


Figure 2 | Efficient and specific gene interference with ubiquitous RNAi. **a**, Efficiency of RNAi-mediated knock-down, as assessed using quantitative RT-PCR on RNA prepared from viable *Act5C-GAL4/UAS-IR* adults and controls. Sixty-four test genes are ranked according to degree of knock down. Data are mean \pm s.e.m. ($n = 2$). **b**, Percentages of positive-control, random and negative-control lines producing lethality or an adult morphology phenotype in combination with *Act5C-GAL4*. The only unexpected phenotype in the negative-control set was the lethality of RNAi against *Dredd*. **c**, Examples of morphological phenotypes resulting from ubiquitous RNAi (middle) compared to a classical loss-of-function mutation (right) and wild type (left). Transformer 2 (*tra2*) RNAi and mutant females anatomically resemble males, including male genitalia and abdominal pigmentation (arrowheads). The eyes are greatly reduced or absent in eyes absent (*eya*) RNAi and mutant males, and Stubble (*Sb*) males have short, stubby bristles on the notum (asterisks indicate the shortened scutellar bristles).

Dicer-2 enhances RNAi potency

For our positive-control set, the RNAi phenotypes sometimes corresponded to the null phenotypes reported for classical mutations, but more often resembled a hypomorphic phenotype. This is consistent with the partial reduction in target mRNA levels generally observed for these lines (Fig. 2a). We wondered whether overexpression of components of the RNAi machinery might enhance such phenotypes. In preliminary experiments in which we overexpressed several different factors involved in RNAi (*dicer-1*, ref. 24; *dicer-2*, ref. 25; *argonaute-1*, ref. 26; *argonaute-2*, ref. 27; *R2D2*, ref. 28; and *tudor-SN*, ref. 29), we found that only *dicer-2* (*Dcr-2*) consistently enhanced the transgenic RNAi effect.

When we retested the same set of 50 lines with both the eye and notum drivers, this time co-expressing *UAS-Dcr-2*, we observed an enhanced RNAi effect with 54% and 43% of the lines, respectively (Fig. 3a, b). We suspected that *UAS-Dcr-2* might also increase off-targeting effects, and so also tested 53 lines (41 genes) from the negative control set (Fig. 2b). In both the eye and notum, the false-positive rate increased by 6% in the presence of *UAS-Dcr-2*. We conclude that co-expression of *UAS-Dcr-2* is likely to be a useful approach to enhance transgenic RNAi effects, but caution that it may also enhance off-targeting effects. We recommend that reliability of *Dcr-2* overexpression be tested carefully for each driver and assay.

Targeting neurons and muscles with RNAi

Conditional RNAi transgenes also allow genes to be inactivated specifically in internal tissues that are difficult or impossible to target by classical genetics. Entire neuronal or muscle networks are prime examples. The complexity of these tissues, and in the case of muscles also their multinuclear structure, renders them essentially inaccessible to conventional mosaic strategies involving mitotic recombination. To test the feasibility of RNAi screens targeted to neurons or muscles, we compiled two complementary sets of positive-control genes: 32 genes known or predicted to act in neurons and 27 genes required in muscles (but neither set exclusive to the respective tissue). Our library contains 49 and 42 transgenic RNAi lines, respectively, for these two gene sets. We also selected an additional 129 lines at random, each representing a different gene (Fig. 4a). The RNAi lines in each of these three sets were crossed in duplicate to either the pan-neuronal driver *elav-GAL4* (ref. 30) or the pan-muscle driver *mef2-GAL4* (ref. 31). On the basis of preliminary experiments, we used *UAS-Dcr-2* with *elav-GAL4* but not with *mef2-GAL4*. Progeny were scored for viability and eight simple behavioural defects (Fig. 4b, c).

A lethal or behavioural phenotype was observed in 82% of the neuronal positive-control set tested with *elav-GAL4* (Fig. 4a, b), and 93% of the muscle positive-control set tested with *mef2-GAL4* (Fig. 4a, c). These results demonstrate that transgenic RNAi is potent in both neurons and muscles, and further indicate that, with the appropriate drivers and assay conditions, false-negative rates can indeed be significantly lower than we estimated from the *Act5C-GAL4* data (Fig. 2b). Hit rates amongst the randomly selected lines were 7% with *elav-GAL4* and 14% with *mef2-GAL4*, with 5% positive with both drivers (Fig. 4a).

Conclusions

Our work provides the resources and proof-of-principle for genome-wide tissue-specific RNAi screens in *Drosophila*. All of these RNAi lines are publicly available (<http://www.vdrc.at>). We estimate that over 60% of these lines trigger potent and specific gene interference, and as many as 90% may be functional in combination with the appropriate drivers, assays, and RNAi-enhancing tools such as *UAS-Dcr-2*. As for RNAi in general, our library is subject to the variable efficiency of gene knock-down and the inherent risk of off-targeting effects. Additionally, because our transgenes are inserted at random sites, some false negatives may be caused by poor transgene expression and some false positives by the misregulation of flanking

endogenous genes. Ultimately, these minor limitations might be overcome by the generation of a second independent library of RNAi transgenes (<http://www.shigen.nig.ac.jp/fly/nigfly>).

Our transgenic RNAi library offers a powerful alternative to classical forward genetic screens. Although mutagenesis screens will continue to be useful, particularly for early embryonic development, transgenic RNAi screens should be especially suitable for later stages and whenever tissue-specific gene disruption is required. Indeed,

transgenic RNAi offers the only practical way to genetically screen certain cell types, such as neurons or muscle. Vast collections of GAL4 driver lines are already available to target RNAi to specific cells²³, and inducible GAL4 systems can be used to further restrict RNAi to a selected stage of the fly's lifespan^{32–34}. Combined with these versatile GAL4 expression systems, our transgenic RNAi library opens up almost limitless possibilities for exploring the genetics of *Drosophila* development, physiology and pathology.

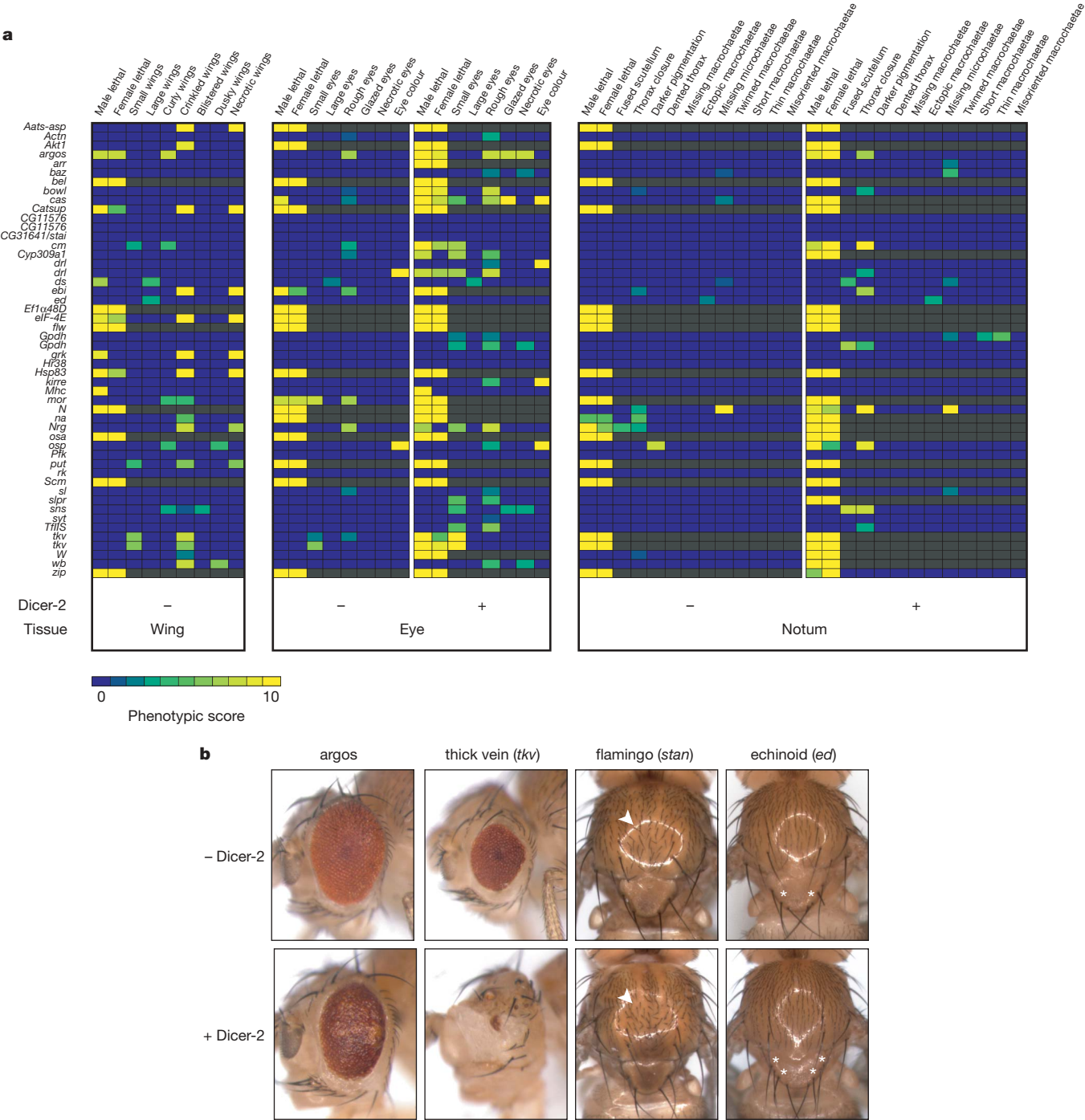


Figure 3 | Tissue-specific RNAi and the enhancing effect of Dicer-2. **a**, Phenotypic data for 50 RNAi lines, representing 46 genes (left), in combination with either *MS1096-GAL4* (wing), *ey-GAL4* and *GMR-GAL4* (eye), or *pnr-GAL4* (notum). For the eye and notum, assays were performed with (+) or without (-) an additional *UAS-Dcr-2* transgene. All phenotypes were scored on a subjective 0–10 scale, with 0 representing no observable defect and 10 the most severe. Phenotypic scores are colour-coded, with grey indicating no data due to lethality. All scores are the average of two

independent replicates. **b**, Examples of RNAi phenotypes in the eye and notum. Wild-type controls are shown in Fig. 2c. RNAi against *argos* results in eye roughening, and RNAi against *tkv* leads to eyes that are both rough and reduced in size. RNAi against *flamingo* (*stan*) leads to a defect in planar cell polarity, evident in the misorientation of microchaetae (arrowheads), whereas RNAi against *ed* results in the formation of ectopic macrochaetae (asterisks). All of the defects shown here are enhanced by co-expression of *UAS-Dcr-2* (bottom row).

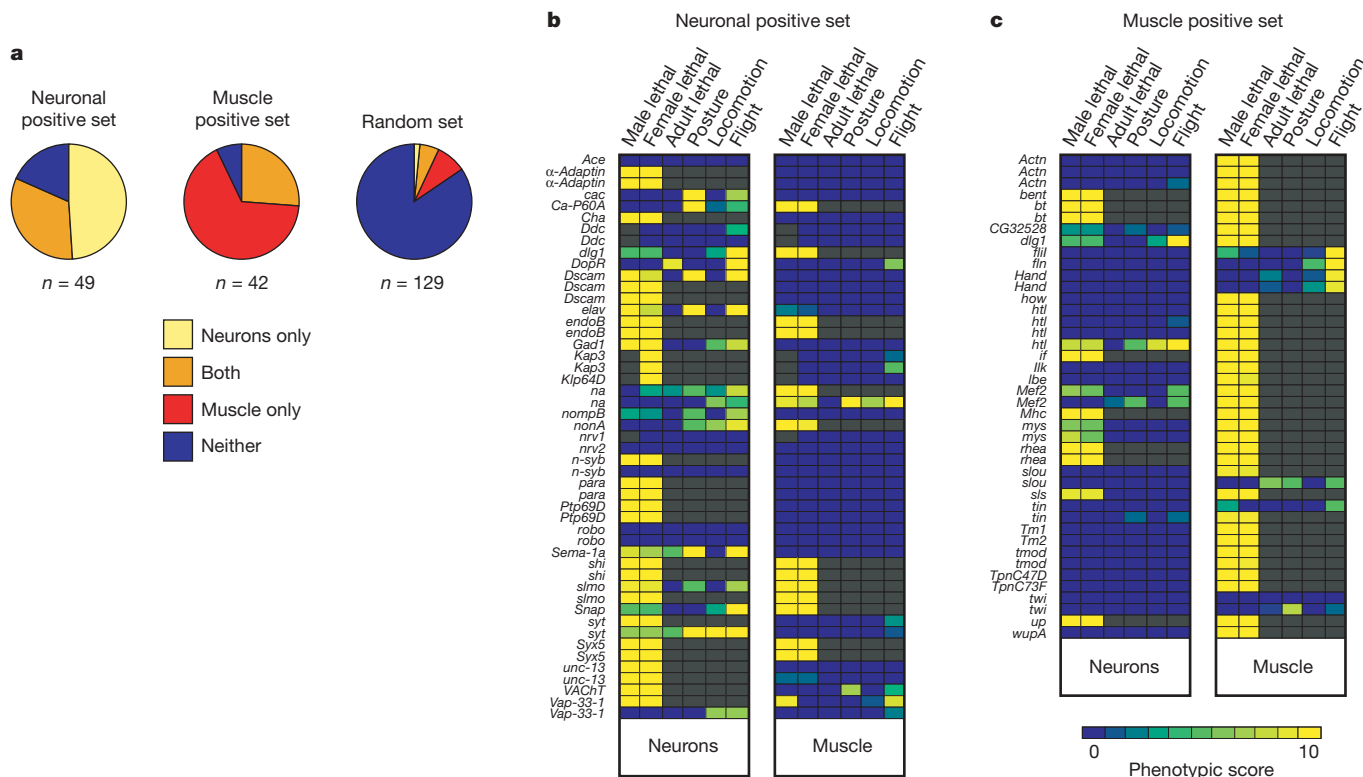


Figure 4 | RNAi in neurons and muscles. **a**, Percentages of neuronal and muscle positive-control lines, and randomly selected lines, that produce either a lethal, postural or behavioural phenotype when expressed with *elav-GAL4* (neurons) or *mef2-GAL4* (muscle). Note that, as anticipated, many of the neuronal positives also produced a phenotype with *mef2-GAL4* (33%), as did the muscle positives with *elav-GAL4* (26%); both sets are enriched but not exclusive for genes required in the respective tissue. **b, c**, Phenotypic data

for 49 lines from the neuronal positive-control set (**b**) and 42 lines from the muscle positive-control set (**c**). Adult lethal indicates lethality by 7 days after eclosion. Data for posture and locomotion are collapsed from 2 and 5 different classes of defect, respectively, scored in the initial analysis. Flight scores indicate the fraction of flies falling immediately to the bottom when tapped into the top of a 1 m × 8 cm cylinder. All values are the average of two independent experiments.

METHODS SUMMARY

UAS-IR constructs. The pMF3 vector is derived from pUAST, but contains 10 rather than 5 tandem copies of the UAS element to enhance expression levels. On the basis of reports that introns enhance nuclear export³⁵, we included a short intron immediately 3' to the inverted repeat. We did not include an intron within the inverted repeat itself, as is often done to improve its stability during cloning^{36–38}, because we did not have any difficulty cloning inverted repeats of less than 500bp and did not observe any enhancement of RNAi efficiency with longer inverted repeats. PCR primers were designed using the Primer3 software³⁹.

Transgenic RNAi lines. Germline transformation was performed using a hyperactive P-element transposase⁴⁰. For verification we extracted genomic DNA from single flies in 96-well plates and performed PCR with one gene-specific primer and a common primer in either the *hsp70* promoter region or the SV40 polyA region (Fig. 1a). Approximately 20% of lines failed the initial verification step; these lines were discarded and the construct re-injected. We believe these are mostly due to the loss of the inverted repeat during propagation in bacteria. To test whether the inverted repeat is stable in *Drosophila*, we verified a set of 192 lines that had been maintained for more than 2 yr (~35 generations) by amplifying both halves of the inverted repeat. In all of these lines, both halves of the inverted repeat were still present, and so we routinely verified transformants by amplifying only one half of the inverted repeat.

Test gene sets. Test genes for quantitative RT-PCR were chosen on the basis of available expression data^{41,42}. We selected genes that were relatively highly expressed, reasoning that they should provide a robust measure of knock-down efficiency, which would, if anything, underestimate the average efficiency. Test sets for functional assays were chosen by manual curation from FlyBase and the literature. For all assays, transgenic RNAi males were crossed in duplicate to virgins carrying the relevant *GAL4* driver and, where appropriate, *UAS-Dcr-2*. Progeny were raised at 25 °C and scored blind to their genotype. Phenotypic categories in Figs 3 and 4 were partly derived by collapsing several distinct defects scored in the initial analysis. In all cases, there was very little phenotypic variation within the progeny of a single cross, with the exception of lethality (for which intermediate values indicate the fraction of animals surviving to the adult stage).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 28 February; accepted 22 May 2007.

1. Fire, A. et al. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391**, 806–811 (1998).
2. Boutros, M. et al. Genome-wide RNAi analysis of growth and viability in *Drosophila* cells. *Science* **303**, 832–835 (2004).
3. Berns, K. et al. A large-scale RNAi screen in human cells identifies new components of the p53 pathway. *Nature* **428**, 431–437 (2004).
4. Paddison, P. J. et al. A resource for large-scale RNA-interference-based screens in mammals. *Nature* **428**, 427–431 (2004).
5. Fraser, A. G. et al. Functional genomic analysis of *C. elegans* chromosome I by systematic RNA interference. *Nature* **408**, 325–330 (2000).
6. Gönczy, P. et al. Functional genomic analysis of cell division in *C. elegans* using RNAi of genes on chromosome III. *Nature* **408**, 331–336 (2000).
7. Kamath, R. S. et al. Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. *Nature* **421**, 231–237 (2003).
8. Sönnichsen, B. et al. Full-genome RNAi profiling of early embryogenesis in *Caenorhabditis elegans*. *Nature* **434**, 462–469 (2005).
9. Reddien, P. W., Bermange, A. L., Murfitt, K. J., Jennings, J. R. & Sanchez Alvarado, A. Identification of genes needed for regeneration, stem cell function, and tissue homeostasis by systematic gene perturbation in planaria. *Dev. Cell* **8**, 635–649 (2005).
10. Newmark, P. A., Reddien, P. W., Cebria, F. & Sanchez Alvarado, A. Ingestion of bacterially expressed double-stranded RNA inhibits gene expression in planarians. *Proc. Natl Acad. Sci. USA* **100** (Suppl 1), 11861–11865 (2003).
11. Van Roessel, P., Hayward, N. M., Barros, C. S. & Brand, A. H. Two-color GFP imaging demonstrates cell-autonomy of GAL4-driven RNA interference in *Drosophila*. *Genesis* **34**, 170–173 (2002).
12. Roignant, J. Y. et al. Absence of transitive and systemic pathways allows cell-specific and isoform-specific RNAi in *Drosophila*. *RNA* **9**, 299–308 (2003).
13. Fortier, E. & Belote, J. M. Temperature-dependent gene silencing by an expressed inverted repeat in *Drosophila*. *Genesis* **26**, 240–244 (2000).
14. Kennerdell, J. R. & Carthew, R. W. Heritable gene silencing in *Drosophila* using double-stranded RNA. *Nature Biotechnol.* **18**, 896–898 (2000).

15. Lam, G. & Thummel, C. S. Inducible expression of double-stranded RNA directs specific genetic interference in *Drosophila*. *Curr. Biol.* **10**, 957–963 (2000).
16. Martinek, S. & Young, M. W. Specific genetic interference with behavioral rhythms in *Drosophila* by expression of inverted repeats. *Genetics* **156**, 1717–1725 (2000).
17. Brand, A. H. & Perrimon, N. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* **118**, 401–415 (1993).
18. Schuler, G. D. Sequence mapping by electronic PCR. *Genome Res.* **7**, 541–550 (1997).
19. Rubin, G. M. & Spradling, A. C. Genetic transformation of *Drosophila* with transposable element vectors. *Science* **218**, 348–353 (1982).
20. Ryder, E. *et al.* The DrosDel collection: a set of P-element insertions for generating custom chromosomal aberrations in *Drosophila melanogaster*. *Genetics* **167**, 797–813 (2004).
21. Kulkarni, M. M. *et al.* Evidence of off-target effects associated with long dsRNAs in *Drosophila melanogaster* cell-based assays. *Nature Methods* **3**, 833–838 (2006).
22. Ma, Y., Creanga, A., Lum, L. & Beachy, P. A. Prevalence of off-target effects in *Drosophila* RNA interference screens. *Nature* **443**, 359–363 (2006).
23. Duffy, J. B. GAL4 system in *Drosophila*: a fly geneticist's Swiss army knife. *Genesis* **34**, 1–15 (2002).
24. Bernstein, E., Caudy, A. A., Hammond, S. M. & Hannon, G. J. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* **409**, 363–366 (2001).
25. Lee, Y. S. *et al.* Distinct roles for *Drosophila* Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways. *Cell* **117**, 69–81 (2004).
26. Williams, R. W. & Rubin, G. M. ARGONAUTE1 is required for efficient RNA interference in *Drosophila* embryos. *Proc. Natl Acad. Sci. USA* **99**, 6889–6894 (2002).
27. Hammond, S. M., Boettcher, S., Caudy, A. A., Kobayashi, R. & Hannon, G. J. Argonaute2, a link between genetic and biochemical analyses of RNAi. *Science* **293**, 1146–1150 (2001).
28. Liu, Q. *et al.* R2D2, a bridge between the initiation and effector steps of the *Drosophila* RNAi pathway. *Science* **301**, 1921–1925 (2003).
29. Caudy, A. A. *et al.* A micrococcal nuclease homologue in RNAi effector complexes. *Nature* **425**, 411–414 (2003).
30. Luo, L., Liao, Y. J., Jan, L. Y. & Jan, Y. N. Distinct morphogenetic functions of similar small GTPases: *Drosophila* Drac1 is involved in axonal outgrowth and myoblast fusion. *Genes Dev.* **8**, 1787–1802 (1994).
31. Ranganayakulu, G., Schulz, R. A. & Olson, E. N. Wingless signaling induces *nautilus* expression in the ventral mesoderm of the *Drosophila* embryo. *Dev. Biol.* **176**, 143–148 (1996).
32. Roman, G., Endo, K., Zong, L. & Davis, R. L. P{Switch}, a system for spatial and temporal control of gene expression in *Drosophila melanogaster*. *Proc. Natl Acad. Sci. USA* **98**, 12602–12607 (2001).
33. Stebbins, M. J. *et al.* Tetracycline-inducible systems for *Drosophila*. *Proc. Natl Acad. Sci. USA* **98**, 10775–10780 (2001).
34. McGuire, S. E., Le, P. T., Osborn, A. J., Matsumoto, K. & Davis, R. L. Spatiotemporal rescue of memory dysfunction in *Drosophila*. *Science* **302**, 1765–1768 (2003).
35. Maniatis, T. & Reed, R. An extensive network of coupling among gene expression machines. *Nature* **416**, 499–506 (2002).
36. Kalidas, S. & Smith, D. P. Novel genomic cDNA hybrids produce effective RNA interference in adult *Drosophila*. *Neuron* **33**, 177–184 (2002).
37. Lee, Y. S. & Carthew, R. W. Making a better RNAi vector for *Drosophila*: use of intron spacers. *Methods* **30**, 322–329 (2003).
38. Reichhart, J. M. *et al.* Splice-activated UAS hairpin vector gives complete RNAi knockout of single or double target transcripts in *Drosophila melanogaster*. *Genesis* **34**, 160–164 (2002).
39. Rozen, S. & Skaletsky, H. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol. Biol.* **132**, 365–386 (2000).
40. Beall, E. L., Mahoney, M. B. & Rio, D. C. Identification and analysis of a hyperactive mutant form of *Drosophila* P-element transposase. *Genetics* **162**, 217–227 (2002).
41. Tomancak, P. *et al.* Systematic determination of patterns of gene expression during *Drosophila* embryogenesis. *Genome. Biol.* **3**, doi:10.1186/gb-2002-3-12-research0088 (2002).
42. Arbeitman, M. N. *et al.* Gene expression during the life cycle of *Drosophila melanogaster*. *Science* **297**, 2270–2275 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank S. Bicker, J. Buch, M. Garstkiewicz, A. Gruber, D. Hofmann Rodrigues, K. Jäger, S. Krüttner, J. Mayerhofer, D. Muggenheimer, E. Muhr, K. Schernhuber, J. Schluder, A. Schmatz, C. Sturtzel, M. Vinzenz and W. Wolfgang for technical assistance, R. Lehmann for the original *hs-hid* lines, J. Mummery-Widmer, M. Yamazaki and J. Knoblich for suggesting the use of *pnr-GAL4*, R. Carthew for initial discussions on the effectiveness of UAS-IR transgenes, and B. Thompson and V. Siegel for helpful comments on the manuscript. F.S. was supported by a long-term postdoctoral fellowship from the Human Frontier Science Program. This work was supported by funds from the Austrian Academy of Sciences (IMBA) and Boehringer Ingelheim GmbH (IMP), and grants from the Austrian Science Fund and the European Union Framework Programme

Author Contributions G.D. established the methodology, and participated in and led the team that constructed the library. K.K. led this team during the finishing stages. D.C. and G.D. performed the bioinformatic analyses, G.D., F.S. and M.F. compiled all the data, which were analysed by G.D., B.J.D. and F.S. K.S. established the enhancing effect of UAS-Dcr-2. B.J.D. conceived and coordinated the project, and wrote the manuscript with input from G.D. and F.S. The remaining authors made major technical contributions to the construction of the library.

Author Information The transgenic RNAi library, including detailed information on each line, is available from the VDRC at (<http://www.vdrc.at>). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to B.J.D. (dickson@imp.ac.at).

METHODS

Primer design. Primers were designed using the Primer3 software³⁹, generally selecting primers of 20–36 nucleotides, a melting temperature (T_m) of 66–70 °C, and a product size of 300–400 bp. These parameters were relaxed if no suitable primers were found. Usually, an *EcoRI* site was added to the left primer and an *XbaI* site to the right primer; if an endogenous recognition site for either enzyme occurred in the predicted product, *BamHI*, *BglII* or *MunI* was used as an alternative. e-PCR⁴³ was performed using a local version of Reverse e-PCR (version 2.3.0 for Windows). The template for e-PCR was either the Release 4.3 genome or transcriptome sequence, according to the template used for the actual PCR. For the transcriptome sequence, if the untranslated regions were unannotated, the protein coding sequence was used and extended by 50 bp of 5' and 150 bp of 3' genomic sequence. We allowed up to 2 mismatches and 2 gaps per primer, but not within the twelve 3' nucleotides, and ± 25 bp size variability compared to the expected product length.

Target predictions. Custom-designed Perl scripts were used to extract all possible 19-mers from all inverted repeat sequences and the Release 4.3 transcriptome, and to search for all perfect matches in either orientation. Target genes for each construct were ranked according to the total number of matches. The primary gene was defined as the top-ranked target (usually hit by 100% of 19-mers, but sometimes fewer owing to changes in the genome annotation). The number of hits on the second-ranked gene (if any) was used for the analysis in Fig. 1c. For calculation of s_{19} scores, on- and off-targets were defined using a cut-off of 80% of the number of matches on the primary target. A custom-designed Java script was used to determine the maximum number of contiguous CAN triplets.

Preparation of *UAS-IR* constructs. Target sequences were amplified by PCR from w^{1118} genomic DNA or mixed stage cDNA in 100 μ l reactions. Fifty microlitres of the reaction was purified by gel filtration (SigmaSpin), digested with *EcoRI*, re-purified by gel filtration, ligated and digested with *XbaI*. The inverted repeats were then separated by gel electrophoresis, purified using NucleoSpin Multi-96 Extract kits (Macherey-Nagel), and ligated into pMF3. All cloning steps were performed manually in 96-well plates. Enzymes other than *EcoRI* and *XbaI* were used in some cases (see above). *Escherichia coli* SURE bacteria were transformed and plated in custom 48-well plates. Two colonies were grown from each transformation, and plasmid DNA extracted using QIAprep 96 Turbo Miniprep kits (Qiagen). The yield was typically 50 μ l eluate with 150–200 ng μ l⁻¹ DNA per construct. Four microlitres of the extracted DNA was digested with *XbaI* and separated by gel electrophoresis to confirm the presence of the inverted repeat. For the positive clones with inserts of the predicted size, 10 μ l of plasmid DNA was prepared for embryo injections by mixing with 3 μ g of $\Delta 2-3$ transposase in a total volume of 12 μ l.

Germline transformation. For each *UAS-IR* construct, ~ 40 dechorionated w^{1118} embryos were aligned on a 24 \times 24 mm glass coverslip, dried for 20 min, covered with 10S Voltaef oil and microinjected using an Eppendorf FemtoJet with a micromanipulator mounted on a Zeiss Axiovert 200 inverted microscope. Borosilicate glass capillaries GC120TF-10 from Harvard Apparatus were pulled on a Sutter P-97 micropipette puller and opened on a Narishige EG-400 micropipette grinder. Injected embryos were placed in a moist chamber at 18 °C for 2 days, and then transferred to fly food vials supplemented with yeast paste and raised at 25 °C. Surviving adults were crossed out to w^{1118} flies, and transformants selected in the progeny using the w^+ marker. We typically obtained at least one transformant in 50–60% of each injection series. Initially, up to 4 lines were established per construct, but typically only up to 2 were retained after verification.

Verification and mapping. We extracted genomic DNA from single flies in 96-well plates. PCR reactions were set up with one specific primer (the right primer used to amplify the inverted repeat fragment) and one common primer (5'-GAGGCGCTTCGTCTACGAGCGAC, located in the *hsp70* promoter region, or in some cases 5'-GCGCTCTAGACGTGTAGTACACTTCGCTACGCAG, located in the SV40 polyA region; see Fig. 1a). Verified lines were then mapped to a specific chromosome by first crossing single transgenic males to virgins obtained from w^{1118}/Y , *hs-hid*; *Sp/CyO* and w^{1118}/Y , *hs-hid*; MKRS/TM2, y^+ stocks, and then crossing transgenic male *CyO* or MKRS progeny to w^{1118} ; *Sp*, *hs-hid/CyO* and w^{1118} ; *Ly*, *hs-hid/TM3 Sb* virgins, respectively. For the latter stocks, the original w^+ marker on each of the *hs-hid* transgenes had previously been removed by imprecise excision mediated by P-element transposase, so as to avoid any possible confusion between this w^+ marker and the w^+ marker on the *UAS-IR* transgene. Once the progeny of these crosses reached the 3rd instar stage, the parents were discarded and the vial heat shocked at 37 °C in a water bath for 1 h to kill progeny carrying the *hs-hid* chromosome. The surviving adults were then scored to determine which chromosome carried the *UAS-IR* transgene, and allowed to mate *inter se* to establish a balanced stock. Homozygous virgins and males were selected in subsequent generations to eliminate balancer

chromosomes, where possible. For insertions on the X chromosome, *UAS-IR* males were crossed to $y w f^-$ virgins obtained from a winscy, *hs-hid/y w f^-* Y stock.

Quantitative RT-PCR. We designed PCR primers (19–23 nucleotides, T_m 60–63 °C) to amplify 150–350 bp fragments from the target transcripts that do not overlap with the original target regions of the *UAS-IR* constructs. Total RNA was extracted from twenty-five 2–4-day-old adult males using Trizol, yielding RNA at $\sim 1.5 \mu$ g μ l⁻¹. RNA (5 μ g) was reverse transcribed using dN₆ random oligos and the SuperScript II Reverse Transcriptase kit (Invitrogen), yielding cDNA at $\sim 1 \mu$ g μ l⁻¹. Quantitative PCR reactions were then performed in 20 μ l reactions using the SYBR Green JumpStart Taq ReadyMix kit (Sigma) and the Mastercycler realplex eppgradientS system (Eppendorf) in 96-well optical plates (Eppendorf). For calibration experiments, PCRs for the test gene and α -tubulin (used as an internal standard) were run on a dilution series of 500, 50, 5 and 0 ng of wild-type cDNA template. After 40 cycles, dissociation analysis was performed to exclude from further experiments those primer pairs that produced primer dimers or amplified non-specific products. Standard curves were calculated for the test genes and α -tubulin⁴⁴, and experiments with correlation coefficients below 0.990 were excluded from further analysis. RNAi flies and controls were then assayed in duplicate by Real-Time PCR for the respective target gene and α -tubulin. The threshold cycle (C_t) values of the target gene were plotted on the wild-type standard curve to calculate target gene levels relative to wild type, and normalized for total cDNA input using α -tubulin levels. Relative knock-down levels of the two samples were averaged.

RNAi machinery transgenes. *UAS-Dcr-1*, *UAS-Dcr-2*, *UAS-AGO1*, *UAS-AGO2*, *UAS-tudor-SN*, and *UAS-R2D2* were prepared by amplifying the corresponding coding regions from genomic DNA and cloning them into the pUAST vector¹⁷, verified by DNA sequencing, and injected into w^{1118} embryos for germline transformation¹⁹. We tested several insertions of each transgene in combination with RNAi transgenes against white (*w*) and sevenless (*sev*). For the experiments reported here, a *UAS-Dcr-2* transgene on the first chromosome was used. Similar enhancement effects were also observed with each of the three other *UAS-Dcr-2* transgene insertions tested.

Test sets. For the *Act5C-GAL4* assays, positive controls were selected by systematically extracting from FlyBase all loss-of-function mutant alleles reported to be associated with a lethal or visible phenotype. The final set of 473 genes was selected at random from an initial list of $\sim 1,400$ genes. The negative set was similarly selected from FlyBase, searching for genes with viable amorphic alleles with no reported phenotype, or in some cases a specific visible phenotype such as eye colour or body pigmentation. Positive controls for the *elav-GAL4* and *mef2-GAL4* assays were manually chosen by browsing FlyBase and the literature, selecting genes reported to function in neurons or muscles, though not necessarily exclusively in either. The random set used in these assays was an arbitrary subset of the random set used for the *Act5C-GAL4* assays.

Phenotypic analyses. For the ubiquitous, wing, eye and notum assays, lethal and morphological phenotypes were scored 3–4 d after eclosion. For the *elav-GAL4* and *mef2-GAL4* assays, viability was scored at eclosion and again at 5–7 d. Flies were aged in groups for 5–7 d before being scored for posture, locomotion, and flight. Males were used for autosomal *UAS-IR* insertions; females for insertions on the X chromosome. Flight assays were performed by dumping 20–30 males into a 1 m \times 8 cm diameter plexiglass tube. Locomotion was assayed by tapping the flies to the bottom of the vial immediately before flight test, and estimating the fraction of flies that were slow or uncoordinated when climbing back up the wall of the vial. Wing posture was scored from the fraction of flies displaying an obvious held-out or erected wing phenotype. For all lethal, morphological and behavioural assays, crosses were scored blind to the genotype and duplicates were always run and scored in different batches on different days. Data presented in heat maps are the average of the two assays. For binary classifications, suitable threshold values were selected for each phenotypic parameter and a line was considered positive only if both assays exceeded the threshold.

43. Schuler, G. D. Sequence mapping by electronic PCR. *Genome Res.* 7, 541–550 (1997).

44. Wong, M. L. & Medrano, J. F. Real-time PCR for mRNA quantitation. *Biotechniques* 39, 75–85 (2005).

Structural basis for transcription elongation by bacterial RNA polymerase

Dmitry G. Vassylyev¹, Marina N. Vassylyeva¹, Anna Perederina¹, Tahir H. Tahirov² & Irina Artsimovitch³

The RNA polymerase elongation complex (EC) is both highly stable and processive, rapidly extending RNA chains for thousands of nucleotides. Understanding the mechanisms of elongation and its regulation requires detailed information about the structural organization of the EC. Here we report the 2.5-Å resolution structure of the *Thermus thermophilus* EC; the structure reveals the post-translocated intermediate with the DNA template in the active site available for pairing with the substrate. DNA strand separation occurs one position downstream of the active site, implying that only one substrate at a time can specifically bind to the EC. The upstream edge of the RNA/DNA hybrid stacks on the β' -subunit 'lid' loop, whereas the first displaced RNA base is trapped within a protein pocket, suggesting a mechanism for RNA displacement. The RNA is threaded through the RNA exit channel, where it adopts a conformation mimicking that of a single strand within a double helix, providing insight into a mechanism for hairpin-dependent pausing and termination.

In all organisms, transcription performed by DNA-dependent RNA polymerases (RNAPs) can be divided into three mechanistically and structurally distinct stages: initiation, elongation and termination. During initiation, RNAP recognizes a promoter, unwinds DNA near the start site (register $i + 1$) and begins RNA synthesis, using NTPs as both a primer and the substrates. Initiation is characterized by multiple rounds of abortive synthesis during which RNAP synthesizes and releases short RNA products. Once RNAP has made an RNA molecule 13–15 nucleotide long, in which 7–9 nucleotides pair with the DNA template strand in an RNA/DNA hybrid, the transcription complex undergoes promoter clearance and makes a transition to the elongation phase.

Elongation is highly processive: the EC is capable of the uninterrupted synthesis of RNA chains thousands of nucleotides long, yet becomes abruptly destabilized at terminators that demarcate the RNA end, in many cases with single-nucleotide precision. The interplay between processive synthesis, transient halting at numerous 'roadblocks' and RNA release depends on the intricate network of interactions between RNAP, the nucleic acid signals and/or auxiliary transcription factors within the EC. High-resolution structural studies of the EC are therefore of central importance for understanding the general mechanisms and basic principles of transcription and the regulation of gene expression in the living cell.

In contrast with initiation complexes, which are vastly different in protein content and complexity, the core enzymes that perform elongation are highly homologous (in both sequence¹ and structure²) in prokaryotic and eukaryotic multisubunit RNAPs. Thus, the structural results and implications obtained on the ECs from either system would be of general interest. Until now there have been no structural data for the bacterial EC, whereas several medium-resolution structures were determined for various configurations of the EC formed by yeast RNAPII^{3–7}. Although these studies yielded important insights into the overall arrangement of the nucleic acids within the EC, the available structures lack many details crucial for understanding the mechanistic aspects of transcript elongation. Moreover, the results and their interpretation provided by different groups were

generally consistent but left several important issues unresolved, including, for example, the point of downstream DNA (dwDNA) strand separation. This further underscores the necessity of obtaining more definitive and highly resolved structural data for the multi-subunit ECs.

We have determined the structure of the *T. thermophilus* RNAP EC (ttEC) at 2.5 Å resolution in which the core enzyme (subunit composition $\alpha_2\beta\beta'\omega$) was bound to a synthetic scaffold containing 14 base pairs (bp) of the dwDNA, 9 bp of the RNA/DNA hybrid and seven single-stranded nucleotides of the displaced RNA transcript (Fig. 1 and Supplementary Fig. 1; see also Methods). This structure—at the highest resolution so far for any multisubunit EC—unambiguously identifies the positions of all the nucleic acid chains at the atomic level and reveals the detailed network of their interactions with the protein side chains, thereby providing a framework for explaining the determinants of the high stability and processivity of the EC. Our data provide several testable implications on several crucial aspects that have remained obscure until now: first, the dwDNA configuration and the mechanism of the DNA strand separation; second, the optimal length of the RNA/DNA hybrid and the mechanism of RNA displacement from the template; third, the mechanism of DNA translocation; and last, the mechanism of the hairpin-dependent pausing and termination.

The core enzyme structure in the EC

Formation of the EC is not accompanied by large alterations of the core RNAP structure: indeed, in the holoenzyme⁸ (which contains an additional protein subunit, σ , that enables promoter-specific transcription initiation) all functionally significant structural elements—the RNA exit channel as well as the binding cavities for the dwDNA and the RNA/DNA hybrid—are already preformed. The last two are, however, partly occupied by the domains of σ that hinder the growth of nascent RNA—thus, the release of most⁹ of the σ -factor contacts with the core is required for the formation of the stable processive EC. This is consistent with the recently reported 'scrunching' of nucleic acids^{10,11} during the transition from initiation to elongation in the

¹Department of Biochemistry and Molecular Genetics, University of Alabama at Birmingham, Schools of Medicine and Dentistry, 402B Kaul Genetics Building, 720 20th Street South, Birmingham, Alabama 35294, USA. ²Eppley Institute for Research in Cancer and Allied Diseases, University of Nebraska Medical Center, Lied Transplant Center, 10737A, 986805 Nebraska Medical Center, Omaha, Nebraska 68198-7696, USA. ³Department of Microbiology, The Ohio State University, 484 West 12th Avenue, Columbus, Ohio 43210, USA.

bacterial multisubunit RNAP, and contrasts with the marked structural refolding that accompanies this transition in the single-subunit T7 RNAP^{12,13}. In the bacterial system, the displacement and eventual release of the σ -factor takes the place of the refolding that is required to expand the active-site cavity and to accommodate the 8–9-bp-long RNA/DNA hybrid in the T7 system.

Nevertheless, the core enzyme conformations in the EC and the holoenzyme are not identical, demonstrating the most significant changes in those structural domains that switch their contacts from the σ -subunit in the holoenzyme to the nucleic acids in the EC. These alterations demarcate the transition from the initiation to the elongation phase and probably contribute to the high stability of the EC, a feature that distinguishes it markedly from the initiation complex. The most notable rearrangements involve the β and β' domains forming the pincers of the crab-claw-like structure of the core enzyme¹⁴ to constitute the main channel accommodating the dwDNA and the RNA/DNA hybrid (Fig. 1 and Supplementary Fig. 2). First, the β' -pincer (the major part of the 'clamp' in the eukaryotic system¹⁵) domain consisting of the amino-terminal α -helical coiled coil (β' CC1; residues 540–581), the major binding site for the σ -subunit, and two loops ('rudder', β' 582–602, and 'lid', β' 525–539)

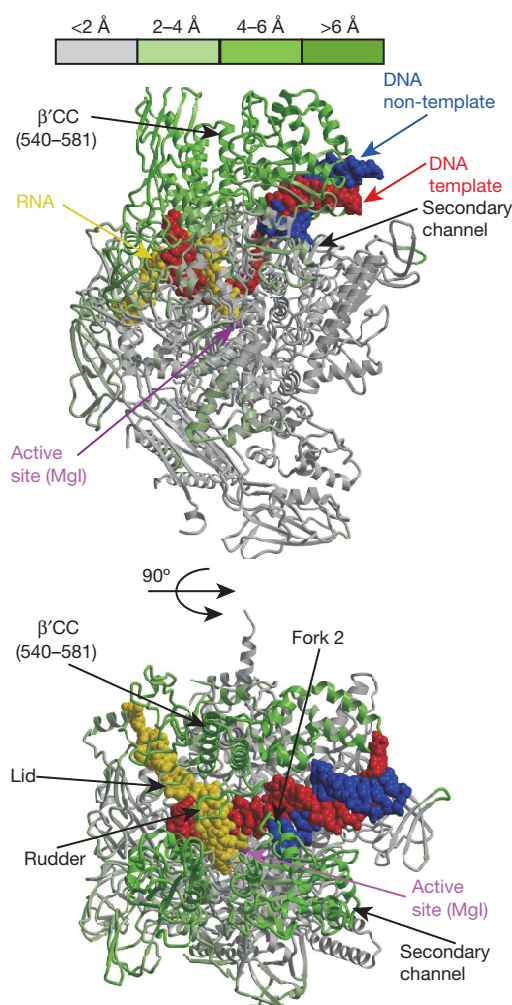


Figure 1 | Overall structure of the ttEC. In the two views shown (the lower panel shows the view through the RNAP main channel), the RNAP core enzyme structure in the ttEC is coloured from grey (less than 2 Å) to green (more than 6 Å) according to the deviation in the positions of the C_{α} atoms of the corresponding core and holoenzyme (PDB ID 2A6H) residues. Except for the variable colour palette used for RNAP in this figure, the same colour scheme (grey, RNAP; red, DNA template; blue, DNA non-template; yellow, RNA; magenta, the high-affinity Mg^{2+} ion, MgI, in the active site) is used in all figures.

158

moves away from its position in the holoenzyme by ~ 5.5 Å, bringing the lid into a stack with the upstream edge of the RNA/DNA hybrid and positioning the rudder right between the dwDNA and the hybrid (see below) (Supplementary Fig. 2a, b). Second, the movement of the coiled coil is accompanied by the corresponding shift of the adjacent β' N-terminal domain (β' 51–499), which places it near the downstream edge of the dwDNA on one side (β' 102–132; β' 470–499) and contributes to the formation of the RNA exit channel on the other (β' 51–87; Supplementary Fig. 2a, c, d). Third, in the β -pincer portion of the claws, the β -domain (β 1–130; β 335–396) that has lost its interactions with the σ -regions 2.4–3.1 (ND2 domain)⁸ moves by ~ 3.0 Å towards the RNA/DNA hybrid, whereas the adjacent β -domain (β 140–332) approaches dwDNA after the ~ 4.0 Å reorientation, thereby closing the dwDNA-binding cleft in the EC (Supplementary Fig. 2a–c). The latter movement is probably attributable to the presence of the dwDNA rather than to the absence of the σ -subunit. Last, the entire β -flap (β 703–830) domain that forms one wall of the RNA exit channel moves slightly (~ 2.2 Å) towards the displaced transcript in the EC (Supplementary Fig. 2d). However, this rearrangement is compensated for by the corresponding repositioning of the β' N-terminal domain (see above) forming another wall of the channel; thus, the size and shape of the RNA exit channel remain nearly unchanged. Overall, in comparison with the holoenzyme, the EC structure undergoes closure of the RNAP claws to reduce the size of the main channel substantially and probably to achieve surface complementarity between the protein and the nucleic acid chains, thereby favouring the high stability and processivity of the EC.

Nucleic acid structure

As observed previously in the structures of T7 RNAP^{13,16,17} and the yeast EC^{4,5}, the template strand in the ttEC forms a sharp kink ($\sim 90^\circ$) at the junction between the dwDNA and the RNA/DNA hybrid (Fig. 1 and Supplementary Fig. 1). The orientation of the nucleic acid duplexes differs substantially from that in the T7 EC (which possesses a distinct architecture) yet closely resembles the eukaryotic complexes, underscoring the structural homology between multisubunit RNAPs (Fig. 2a).

The nascent seven-nucleotide single-stranded RNA is threaded towards the surface through the previously proposed RNA exit channel¹⁸. In the channel, six out of seven RNA bases stack on each other, whereas the phosphate backbone conformation mimics that in the A-form nucleic acid duplex (Fig. 2b); only a few direct and several water-mediated hydrogen bonds are formed between the protein and the RNA nucleotides (Fig. 3), suggesting that the RNA conformation is maintained mostly through the self-stabilizing base-stacking contacts and/or solvent-mediated interactions. Several acidic side chains and carbonyl main-chain oxygens approach the RNA backbone, indicative of a possible involvement of the Mg^{2+} ions in mediating these interactions.

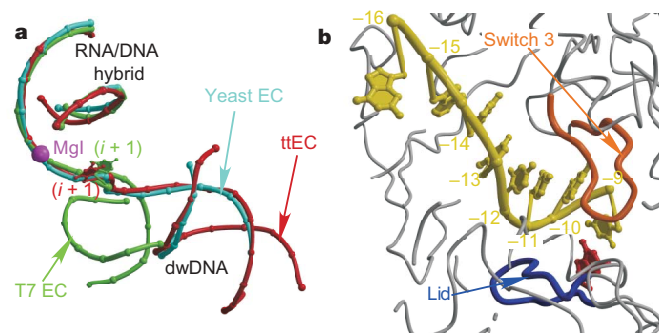


Figure 2 | Nucleic acid structure in the ttEC. **a**, Superposition of the nucleic acid phosphate backbones in the ttEC (red), yeast EC (cyan)⁴ and T7 EC (green)¹⁶. **b**, Conformation of the displaced RNA nucleotide in the ttEC.

Protein–nucleic-acid interactions

The dwDNA adopts a B-form conformation that is substantially distorted only at registers +2 and +3, immediately downstream of the unpaired acceptor DNA template (register $i + 1$) where the kink occurs. A similar distortion was also observed in the eukaryotic EC structure⁴, further underscoring the similarity between the two systems. Although all 13 bp of the dwDNA are buried within the closed channel formed by the β - and β' -pinchers, there are very few direct polar interactions with the protein at the downstream dwDNA edge (Fig. 3), suggesting that the DNA conformation is maintained through self-stabilizing base pairing, whereas the protein stabilizes the duplex orientation through long-distance electrostatic and van der Waals interactions. The same pattern was observed in the T7 EC^{13,16}, implying that dwDNA mobility is a general feature that affords high processivity during elongation. The major contacts of RNAP to the dwDNA between registers +2 and +5 probably stabilize the kink in the nucleic acid and help to position the acceptor template base properly in the active site (Fig. 3). Particularly noteworthy is a β' -subunit three-helix bundle (β' 101–131) that forms several polar and van der Waals interactions with the dwDNA phosphates (Fig. 4a and Supplementary Fig. 3a). Val β' 108 located in the loop between the first and second helices inserts into the minor groove (Fig. 4a) and may prevent translational sliding of the nucleic acids while inducing the screw-like motion during DNA translocation. His 772 probably has a similar function in the T7 EC (Fig. 4b). In yeast RNAPII this region corresponds to the N terminus of the clamp head, which (according to the structural alignment) contains a uniform helix that superimposes well on the first and second helices of the bacterial bundle but lacks the Val β' 108-loop insertion (Supplementary Fig. 4a). Inspection of the yeast EC structure⁴ revealed another structural

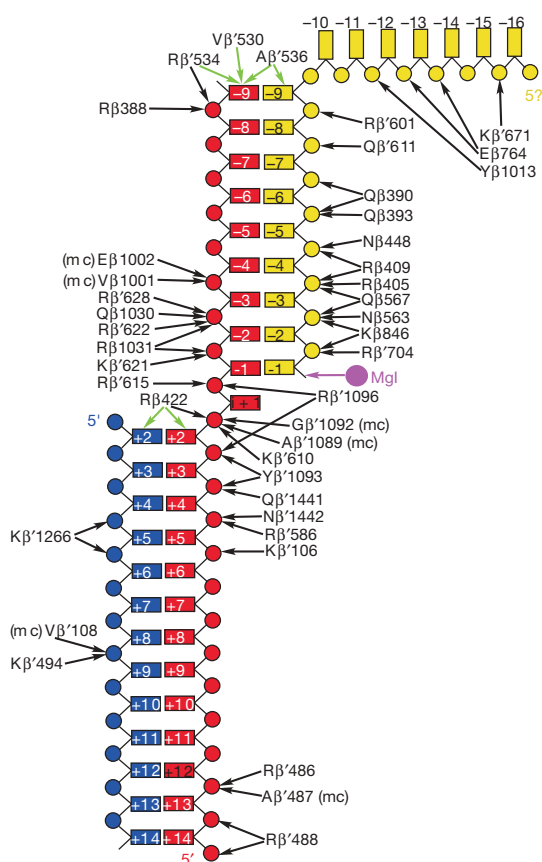


Figure 3 | Schematic drawing of the protein/nucleic acid contacts. Polar and stacking interactions are shown as black and green arrows, respectively. The ($i + 1$) non-template nucleotide was not resolved in the electron density map and is therefore not shown.

element at the very carboxy terminus of the cleft domain, which is represented by a uniform α -helix in the bacterial enzyme but is interrupted by insertions of two Rpb1 residues, Arg 1386 and His 1387, whose side chains are positioned in the dwDNA minor groove mimicking the Val β' 108 loop (Fig. 4c and Supplementary Fig. 4b). The exact location of the loop insertion into the minor groove is, however, different in each complex: in yeast and T7 ECs it occurs closer to the upstream edge of the dwDNA than to the ttEC.

The RNA/DNA hybrid is tightly packed in the active-site cavity, where its phosphate backbone forms multiple polar and van der Waals interactions with the conserved protein residues (Fig. 3; see also Supplementary Table 2 in ref. 19). In this 'mixed' mode of binding (also observed in the T7 EC^{13,16}) the van der Waals contacts may have two essential functions. First, repulsion between the hydrophobic side chains and charged phosphate groups may weaken hydrogen bonds that the latter form with the basic residues, thereby avoiding excessively strong binding of the nucleic acids that could reduce the EC processivity. Second, they may act as 'shape sensors' that monitor the hybrid configuration to provide feedback in response to mismatches and/or occasional dNMP incorporations, thereby contributing to the fidelity of transcription.

The rudder loop is positioned between the RNA/DNA hybrid and the dwDNA (Fig. 4d), suggesting that although it is unlikely to function in RNA displacement and/or separation of the upstream hybrid

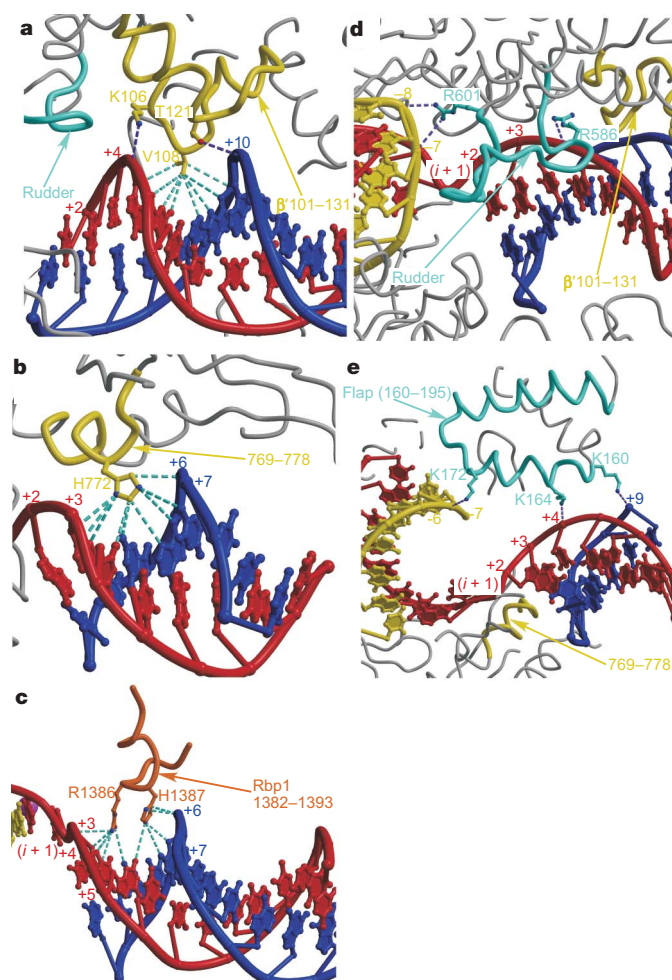


Figure 4 | Protein–nucleic acid interactions in the ttEC structure.

a–c, Protein segments inserted in the minor groove of the dwDNA in the ttEC (**a**), T7 EC (**b**) and eukaryotic EC (**c**). **d**, **e**, Rudder loop (**d**) and flap domain (**e**) demarcate the RNA/DNA hybrid and dwDNA-binding sites in the ttEC and T7 EC, respectively. The van der Waals and polar interactions are shown by cyan and purple dashed lines, respectively.

strand^{18,20}, it contributes to the overall stability of the transcription bubble, a function consistent with the biochemical data²¹. Indeed, the rudder demarcates the hybrid and the dwDNA, while stabilizing their position and conformation through the direct contacts; two Arg residues (β' 586 and β' 601) located at its N and C termini interact directly with the dwDNA phosphates and the hybrid, respectively (Fig. 4d). The rudder interacts with the upstream edge of the RNA/DNA hybrid (registers $-7/-8$); thus, the contacts bridging the hybrid and the dwDNA may account for the higher stability of the EC than the early initiation complex. In contrast, the proper positioning of the rudder and the setting-up of these bridging interactions may help to achieve the successful transition to elongation. The flap subdomain in T7 RNAP (which is characteristic of the EC but is missing from the initiation complex^{12,13,16,22}) also bridges the hybrid and the dwDNA and is probably a functional counterpart of the rudder²³ (Fig. 4e).

dwDNA and RNA/DNA hybrid strand separation

During elongation, RNAP performs thousands of nucleotide addition cycles. Each cycle must culminate in forward translocation by one nucleotide to allow for the incorporation of the next substrate; this step entails the separation of 1 bp of the dwDNA accompanied by the displacement of one nucleotide of the nascent RNA from the DNA template at the upstream edge of the RNA/DNA hybrid and subsequent annealing of the upstream DNA duplex. The mechanisms of the dwDNA and the RNA/DNA hybrid strand separation have been largely explained for the single-subunit enzymes^{12,13}, but remain poorly understood for the multisubunit RNAPs.

Two models based on the medium-resolution structures of the yeast ECs^{3,6} differ with regard to the position of the melting of the dwDNA: in one model, strand separation occurs deep inside the dwDNA-binding channel (register $+5$)⁶, whereas in another the dwDNA duplex is maintained at least up to the $+3$ register⁴. However, it is difficult to compare these models because the scaffold used in one work lacked the non-template strand up to the $+4$ register⁶, whereas the mismatch was present in the dwDNA at register $+2$ in the other⁴. In the ttEC structure, in which we used the dwDNA that is fully complementary along its entire length (14 bp), the duplex is maintained up to and including the $+2$ register, with only one base pair open at register $+1$ (Supplementary Fig. 1), making the acceptor template in the active site available for base pairing with the incoming substrate. This observation is fully supported by a large body of work on the bacterial transcription complexes (see ref. 18 and references therein) and thus indicates that only one substrate at a time can bind specifically to the EC (see ref. 19 for additional evidence), while arguing against 'multisubstrate' models^{24–26}. It is worth mentioning that our scaffold lacks the non-template strand upstream of the insertion site; the consequent lack of the upstream DNA duplex may indirectly affect the dwDNA strand separation. In the bacterial EC, the N terminus (β 418–424) of loop 2 of the β fork (fork 2, β 413–451; Fig. 1 and Supplementary Fig. 3b) acts as a steric barrier to the passage of the dwDNA towards the active site, probably having a central function in facilitating strand separation. Moreover, Arg β 422 at the tip of this loop forms a hydrogen bond with the acceptor template phosphate and stacks on the $+2$ base pair (Fig. 5a); similarly, in T7 RNAP Phe 664 stacks on the $+2$ template base, and the main chain amino group of the neighbouring Gly 645 is hydrogen-bonded to the acceptor template phosphate¹² (Fig. 5b). It is noteworthy that the F644A and G645A mutants are among four T7 RNAP variants known to show a significant increase in the incorporation of mismatched bases²⁷, suggesting that these interactions may be coupled to substrate loading and selection¹². This analogy suggests, first, that Arg β 422 may be a fidelity determinant in the bacterial RNAP, and second, that base pairing at register $+2$ might be essential for the proper positioning of the acceptor DNA base in the active site, which in turn may affect recognition and binding of the cognate substrate. Although poor sequence alignment and

conformational differences preclude unambiguous identification of the Arg β 422 counterpart in eukaryotic fork 2, two basic residues that are spatially close to Arg β 422 in the yeast structure (Arg 504 and Lys 507)⁴ may have a similar function. In contrast, although the location of fork 2 in one of the eukaryotic ECs is similar, suggesting the same mechanism and position of separation of the dwDNA strand⁴, it adopts different conformations in other complexes⁵ and may disrupt the $+5$ base pair. Obtaining an unambiguous answer to this question would require a structural analysis of the eukaryotic ECs assembled on a nucleic acid scaffold containing a complete dwDNA duplex as in the present study.

The ttEC structure shows that RNAP can readily accommodate 9 bp of the RNA/DNA hybrid; its upstream edge (register -9) interacts with the lid loop that was previously proposed to facilitate the RNA displacement in eukaryotic RNAP³ (Fig. 5c). In contrast to the yeast enzyme, however, in which this loop forms no direct contacts with the nucleic acids, in the ttEC several residues (conserved between bacteria and eukaryotes) from the lid stack onto the upstream base pair of the hybrid, mimicking the base-stacking interactions within the nucleic acid duplex (Fig. 5c and Supplementary Fig. 5). The lid therefore probably sterically blocks further growth of the hybrid, facilitating strand separation while simultaneously stabilizing the upstream base pair. Given the likely flexibility of this protruding structural segment, its potential reorientations could accommodate the 8-bp or 10-bp RNA/DNA hybrids. The lid may thus serve as an important regulatory element in stabilizing various non-standard hybrid configurations that form during pausing, backtracking or termination, for example.

Although deletion of the lid does indeed result in the formation of persistent RNA/DNA hybrids on single-stranded templates, it seems dispensable in the presence of the non-template strand, suggesting that reannealing of the upstream DNA is crucial for displacement of the RNA^{28,29}. In this respect, the trapping of the first displaced RNA base in the hydrophobic pocket formed by the β 1012–1022 loop (switch 3) is of particular interest (Fig. 5d). Indeed, if competition

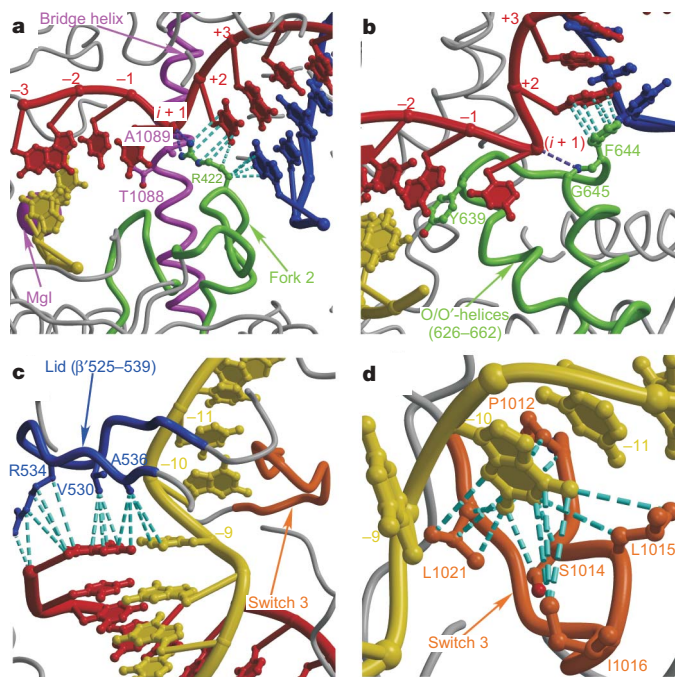


Figure 5 | dwDNA and RNA/DNA hybrid strand separation. **a, b**, Structural elements interacting with the $+2$ dwDNA base pair in the ttEC (**a**) and T7 EC (**b**). **c**, The lid loop stacks on the upstream RNA/DNA base pair in the ttEC. **d**, The first displaced RNA base is trapped in the protein pocket formed by the displacement loop in the ttEC. The van der Waals and polar interactions are shown by cyan and blue dashed lines, respectively.

is assumed between the displaced RNA and the non-template strand nucleotides for the template strand, the switch 3 loop harbouring the RNA base might have an essential function in the 'DNA-dependent' mechanism of the RNA separation.

Active site and implications for translocation

In the post-translocated tTEC, the acceptor template base ($i + 1$) resides in the active site, where it is sandwiched between the -1 template nucleotide and the bridge helix, mimicking the duplex configuration (Fig. 5a). As a result, its base pairing with the substrate may position the substrate phosphates near the active-site residues, in stark contrast to the post-translocated state of the T7 EC, in which the acceptor template base is displaced by the side chain of Tyr 639 (Fig. 5b), 'forcing' the initial substrate binding into the inactive, 'preinsertion' site located far from the active centre¹² and thus necessitating significant structural rearrangements to deliver the substrate to the active, 'insertion' position. Even though the transition to the 'insertion' state in the bacterial RNAP apparently does not call for such large changes, the structures of the substrate-bound tTECs¹⁹ argue for the conceptually similar two-step mechanism: NTP first binds to the preinsertion configuration followed by isomerization to the closed, insertion state after the substrate-induced folding of the trigger loop, which is disordered in the ligand-free tTEC.

Although the protein/nucleic acid interactions formed at the upstream and downstream edges of the hybrid and the dwDNA, respectively, as well as those made with the first displaced RNA nucleotide and the acceptor template, potentially stabilize the post-translocated tTEC state, the same interactions are probably present in the pre-translocated complex, suggesting that these two states are equally stable in the absence of the substrate. Our analysis failed to reveal structural elements (for example, a 'ratchet') that would preclude the backward DNA translocation in the post-translocated tTEC. In particular, the bridge helix, whose local distortion⁸ could potentially freeze the post-translocated state, is uniform in the tTEC. The tTEC structure therefore supports a 'brownian-ratchet' mechanism, in which translocation is driven by thermal motions giving rise to equilibrium between the post-translocated and pre-translocated states of the substrate-free ECs, whereas binding of the cognate substrate would favour the post-translocated state³⁰. In contrast, structures of the T7 RNAP EC reveal markedly distinct orientations of the template in the register ($i + 1$) in the post-translocated and pre-translocated conformations¹³, prompting a hypothesis that the Tyr 639 side chain replacing the DNA template in the active site acts as a 'ratchet' stabilizing the post-translocated state and blocking reverse translocation.

Implications for pausing and termination

Gene expression in bacteria is frequently controlled by nascent RNA hairpins, which, depending on the context, can slow RNAP down at a pause site to await a regulatory input or trigger RNA release at a terminator. The principal difference between the pause and the terminator hairpins is the register at which the RNA duplex begins, namely ($-11/-10$) and (less than -8), respectively^{31,32}. The 9 bp of the RNA/DNA hybrid and the duplex-like conformation of the displaced RNA observed in the tTEC suggest that the pause hairpin could be formed within the exit channel without major alterations of the enzyme structure. Indeed, modelling demonstrates that the RNA exit channel can accommodate the A-form 5-bp hairpin stem after subtle rotations of the experimental displaced RNA strand ($\sim 22^\circ$) used as a template for the modelling, and the flexible flap-tip helix ($\sim 30^\circ$; $\beta 761-785$; Supplementary Fig. 6). The eukaryotic RNA exit channel that lacks the flap-tip helix and seems more open could accommodate nascent hairpins even more readily. In the model, the first displaced RNA base -10 is buried in the protein pocket, retaining its position observed in the experimental structure, and is not available for base pairing. In an alternative model the hairpin-induced local alterations of the lid and/or switch 3 loops may result in

the formation of an additional RNA/DNA hybrid base pair that would increase the stability of the paused complex, while impeding the translocation and/or RNA displacement and prolonging the pause.

Modelling of the termination hairpin seemed impossible in view of the predicted significant rearrangements of the RNAP structure after hairpin formation at register -8 that triggers forward translocation and RNA release³³ or EC inactivation and dissociation³⁴. However, our modelling attempts suggest that these rearrangements would probably involve those structural elements whose orientation is affected by the σ binding in the holoenzyme. Given that both the initiation and termination complexes are highly unstable, it is possible that the termination hairpin would destabilize the complex in a fashion similar to that of the σ -factor. For example, the hairpin might displace the lid/coiled-coil/rudder domain in a σ -like manner (see above), thereby disrupting the rudder-nucleic-acid interactions that affect the EC stability²¹.

Concluding remarks

This study provides the first detailed view of the bacterial EC structure and identifies the determinants of the complex stability, processivity, and response to the regulatory nucleic acid signals, which in turn suggest the mechanisms for separation of the nucleic acid strand, DNA translocation, and transitions from the initiation to the elongation stage and finally to the termination stage of transcription. Most remarkably, it allows us to draw numerous parallels between single-unit and multisubunit enzymes and argue that these RNAPs use analogous structural mechanisms of nucleic acid binding and processing, because many fundamental features (such as duplex separation, RNA displacement and active-site architecture) seem strikingly similar at the molecular level. This conclusion is perhaps not surprising in view of the ancient evolutionary origin of the RNAPs and the similarities in response of the extant enzymes from phages, bacteria and eukaryotes to at least a subset of regulatory signals, yet it remains to be widely appreciated in the field. Indeed, scores of adaptive and accidental evolutionary innovations found in prokaryotes and eukaryotes obscure the profound congruity of the basic transcriptional mechanisms in all life kingdoms. However, the progress in understanding the differences in organization of transcriptional machinery between organisms and the impact they have on the regulation of gene expression depends on an understanding of the elemental similarity of the catalytic mechanisms and their structural framework. This work has fitted into place a few more pieces of the transcription puzzle, while highlighting the need for further advances on two fronts: first, continual improvement in the structural data to the near-ångström resolution required to explain the basic catalytic mechanism, and second, a progressive build-up of the crystallized complexes to include the non-template DNA strand, nascent RNA hairpins, and regulatory factors.

METHODS SUMMARY

The structure was determined by the molecular replacement technique³⁵ by using the core enzyme portion of the *T. thermophilus* holoenzyme structure⁸ (PDB entry 1IW7) as a search model (Supplementary Table 1). Application of the zonal scaling to the experimental diffraction data greatly improved the R_{free} and omit electron density (Supplementary Figs 7 and 8 and Supplementary Table 2) and permitted model building of all nucleic acid components of the EC. The structure was refined to a final R -factor/ R_{free} of 0.238/0.267 at 2.5 Å resolution (Supplementary Table 3).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 28 January; accepted 11 May 2007.

Published online 20 June 2007.

1. Archambault, J. & Friesen, J. D. Genetics of eukaryotic RNA polymerases I, II, and III. *Microbiol. Rev.* **57**, 703–724 (1993).
2. Cramer, P. Multisubunit RNA polymerases. *Curr. Opin. Struct. Biol.* **12**, 89–97 (2002).

3. Gnatt, A. L., Cramer, P., Fu, J., Bushnell, D. A. & Kornberg, R. D. Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 Å resolution. *Science* **292**, 1876–1882 (2001).
4. Kettenberger, H., Armache, K. J. & Cramer, P. Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIS. *Mol. Cell* **16**, 955–965 (2004).
5. Wang, D., Bushnell, D. A., Westover, K. D., Kaplan, C. D. & Kornberg, R. D. Structural basis of transcription: role of the trigger loop in substrate specificity and catalysis. *Cell* **127**, 941–954 (2006).
6. Westover, K. D., Bushnell, D. A. & Kornberg, R. D. Structural basis of transcription: nucleotide selection by rotation in the RNA polymerase II active center. *Cell* **119**, 481–489 (2004).
7. Westover, K. D., Bushnell, D. A. & Kornberg, R. D. Structural basis of transcription: separation of RNA from DNA by RNA polymerase II. *Science* **303**, 1014–1016 (2004).
8. Vassylyev, D. G. *et al.* Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 Å resolution. *Nature* **417**, 712–719 (2002).
9. Mooney, R. A., Darst, S. A. & Landick, R. Sigma and RNA polymerase: an on-again, off-again relationship? *Mol. Cell* **20**, 335–345 (2005).
10. Kapanidis, A. N. *et al.* Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism. *Science* **314**, 1144–1147 (2006).
11. Revyakin, A., Liu, C., Ebright, R. H. & Strick, T. R. Abortive initiation and productive initiation by RNA polymerase involve DNA scrunching. *Science* **314**, 1139–1143 (2006).
12. Temiakov, D. *et al.* Structural basis for substrate selection by T7 RNA polymerase. *Cell* **116**, 381–391 (2004).
13. Yin, Y. W. & Steitz, T. A. The structural mechanism of translocation and helicase activity in T7 RNA polymerase. *Cell* **116**, 393–404 (2004).
14. Zhang, G. *et al.* Crystal structure of *Thermus aquaticus* core RNA polymerase at 3.3 Å resolution. *Cell* **98**, 811–824 (1999).
15. Cramer, P., Bushnell, D. A. & Kornberg, R. D. Structural basis of transcription: RNA polymerase II at 2.8 Å resolution. *Science* **292**, 1863–1876 (2001).
16. Tahir, T. H. *et al.* Structure of a T7 RNA polymerase elongation complex at 2.9 Å resolution. *Nature* **420**, 43–50 (2002).
17. Yin, Y. W. & Steitz, T. A. Structural basis for the transition from initiation to elongation transcription in T7 RNA polymerase. *Science* **298**, 1387–1395 (2002).
18. Korzheva, N. *et al.* A structural model of transcription elongation. *Science* **289**, 619–625 (2000).
19. Vassylyev, D. G. *et al.* Structural basis for substrate loading in bacterial RNA polymerase. *Nature* doi:10.1038/nature05931 (this issue).
20. Naryshkin, N., Revyakin, A., Kim, Y., Mekler, V. & Ebright, R. H. Structural organization of the RNA polymerase-promoter open complex. *Cell* **101**, 601–611 (2000).
21. Kuznedelov, K., Korzheva, N., Mustaev, A. & Severinov, K. Structure-based analysis of RNA polymerase function: the largest subunit's rudder contributes critically to elongation complex stability and is not involved in the maintenance of RNA–DNA hybrid length. *EMBO J.* **21**, 1369–1378 (2002).
22. Cheetham, G. M. & Steitz, T. A. Structure of a transcribing T7 RNA polymerase initiation complex. *Science* **286**, 2305–2309 (1999).
23. Jiang, M., Ma, N., Vassylyev, D. G. & McAllister, W. T. RNA displacement and resolution of the transcription bubble during transcription by T7 RNA polymerase. *Mol. Cell* **15**, 777–788 (2004).
24. Foster, J. E., Holmes, S. F. & Erie, D. A. Allosteric binding of nucleoside triphosphates to RNA polymerase regulates transcription elongation. *Cell* **106**, 243–252 (2001).
25. Gong, X. Q., Zhang, C., Feig, M. & Burton, Z. F. Dynamic error correction and regulation of downstream bubble opening by human RNA polymerase II. *Mol. Cell* **18**, 461–470 (2005).
26. Landick, R. NTP-entry routes in multi-subunit RNA polymerases. *Trends Biochem. Sci.* **30**, 651–654 (2005).
27. Huang, J., Briebe, L. G. & Sousa, R. Misincorporation by wild-type and mutant T7 RNA polymerases: identification of interactions that reduce misincorporation rates by stabilizing the catalytically incompetent open conformation. *Biochemistry* **39**, 11571–11580 (2000).
28. Naryshkina, T., Kuznedelov, K. & Severinov, K. The role of the largest RNA polymerase subunit lid element in preventing the formation of extended RNA–DNA hybrid. *J. Mol. Biol.* **361**, 634–643 (2006).
29. Touloukhonov, I. & Landick, R. The role of the lid element in transcription by *E. coli* RNA polymerase. *J. Mol. Biol.* **361**, 644–658 (2006).
30. Abbondanzieri, E. A., Greenleaf, W. J., Shaevitz, J. W., Landick, R. & Block, S. M. Direct observation of base-pair stepping by RNA polymerase. *Nature* **438**, 460–465 (2005).
31. Artsimovitch, I. & Landick, R. Interaction of a nascent RNA structure with RNA polymerase is required for hairpin-dependent transcriptional pausing but not for transcript release. *Genes Dev.* **12**, 3110–3122 (1998).
32. Yarnell, W. S. & Roberts, J. W. Mechanism of intrinsic transcription termination and antitermination. *Science* **284**, 611–615 (1999).
33. Santangelo, T. J. & Roberts, J. W. Forward translocation is the natural pathway of RNA release at an intrinsic terminator. *Mol. Cell* **14**, 117–126 (2004).
34. Gusarov, I. & Nudler, E. The mechanism of intrinsic transcription termination. *Mol. Cell* **3**, 495–504 (1999).
35. Navaza, J. Implementation of molecular replacement in AMoRe. *Acta Crystallogr. D Biol. Crystallogr.* **57**, 1367–1372 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank D. Temiakov and M. Anikin for assistance in crystallization at the initial stage of the project. We are grateful to R. Landick for helpful discussions and critical reading of the manuscript. Use of the Advanced Photon Source was supported by the US Department of Energy, Office of Energy Research. This work was supported by NIH grants to D.G.V. and I.A.

Author Contributions D.G.V. determined and analysed the structure and guided the project. A.P. purified RNAP. M.N.V. and A.P. performed crystallization and data collection. T.H.T. assisted with data collection and analysis. I.A. contributed scaffold design and analysis. D.G.V. and I.A. jointly wrote the manuscript.

Author Information The atomic coordinates are deposited in the Protein Data Bank under accession number 2O5I. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to D.G.V. (e-mail: dmitry@uab.edu).

METHODS

Crystallization and data collection. Initially, the two nucleic acid scaffolds were designed for co-crystallization with the *T. thermophilus* core enzyme that differed only in the length (8 and 9 bp) of the RNA/DNA hybrid^{36,37}. Surprisingly, the preliminary biochemical characterization of the ECs assembled on these scaffolds (that might be designated as the EC8 and EC9) revealed that whereas the EC9 was resistant, the EC8 seemed highly sensitive to pyrophosphorolysis, suggesting that the EC9 and EC8 are preferentially in the post-translocated and pre-translocated states, respectively. Consistent with this interpretation, crystallization of these complexes resulted in the crystals obtained under essentially distinct crystallization conditions, belonged to the different space groups (tetragonal and hexagonal for the EC9 and EC8, respectively) and were not cross-reproducible^{36,37}. Whereas the EC8 hexagonal (presumably pre-translocated) crystals diffracted to only 7 Å resolution, the very first relatively small EC9 tetragonal crystals diffracted beyond 4 Å and we therefore concentrated on the improvement of these promising crystals.

The optimized crystallization conditions for the EC9 (the well solution contained 30% 2-methyl-2,4-pentandiol, 0.1 M sodium cacodylate pH 6.5 and 0.2 M magnesium acetate) resulted in the large bipyramidal crystals that belonged to the space group $P4_12_1$ with unit cell dimensions $a = b = 156.21$ Å, $c = 499.23$ Å. In spite of their large size, most of the crystals exhibited poor diffraction quality and high mosaicity. After numerous trials, we were able to select only one well-diffracting crystal, from which the preliminary high-resolution data were collected at synchrotron beam line BL6, Photon Factory, Japan^{36,37}. Although the data allowed us to solve the structure by molecular replacement and indicated the presence of the nucleic acids in the complex, the electron density (ED) for the nucleotides was of the overall modest quality and was especially poor for the upstream edge of the RNA/DNA hybrid, the downstream edge of the dwDNA and for the displaced RNA transcript. In particular, the ED corresponding to the two template nucleotides at the upstream edge of the hybrid was nearly absent. We presumed that although the ECs showed high stability during the relatively short biochemical experiments, they might be still affected by the oscillation between the post-translocated and pre-translocated states that would permit the excision of the 3'-RNA nucleotides during crystallization lasting several weeks, thereby resulting in a mixture of the different complexes coexisting in the crystal. The analysis of the nucleic acids from the crystals that were growing for about three months revealed that this hypothesis was correct (Supplementary Fig. 9). To resolve this problem, a new scaffold was designed, in which the two 3'-RNA nucleotides were linked by phosphorothioate bonds to inhibit nucleolytic cleavage, and the G/C content in the hybrid and the dwDNA was increased to achieve better stability of the complexes and presumably prevent backtracking. Indeed, the new scaffold seemed substantially more stable in the lengthy crystallization experiments (Supplementary Fig. 9), which resulted in faster-growing and better-diffracting crystals. Though the percentage of crystals diffracting to high resolution was substantially increased, all of them were still characterized by a high, anisotropic mosaicity that, in combination with a quite weak diffraction power, seriously affected the quality of the processed data. To improve the mosaicity, we applied an annealing procedure to the crystals. After multiple trials we found that freezing the crystals in the cryo-stream followed by annealing and soaking them in the cryoprotectant for ~1–2 min greatly improved the mosaicity for some crystals on the second cycle of freezing. This allowed us finally to select a few well-diffracting crystals with a reasonably low mosaicity (0.5–0.7°), from which the best sample was used for high-resolution data collection at the SERCAT beam line (APS, Argonne, USA); the structure was refined by using the twinning option of the CNS program³⁸ alternating with manual model building³⁹ to a final R -factor/ R_{free} of 0.238/0.267 at 2.5 Å resolution (Supplementary Table 3).

All structural figures were prepared with the programs Molscript, Bobscript and Raster3D^{40–42}.

Zonal scaling: the key for structure solution. Molecular replacement³⁵ applied to the EC data provided a clear single solution for both rotation and translation functions using the core enzyme extracted from the previously determined structure of the *T. thermophilus* holoenzyme⁸ (PDB entry 1IW7) as a search model (Supplementary Table 1). Surprisingly, however, the rigid body and positional refinements performed against the original diffraction data were highly unstable and did not yield a decisive decrease in R -factor from the initial value of ~49.8% at 2.5 Å resolution (Supplementary Table 2) providing, in particular, the abnormally high R -factors in the lowest resolution shell. As a result, the $(2F_{\text{obs}} - F_{\text{calc}})$ ED map was characterized by a largely smeared ED in the protein region that was lacking the atomic details, demonstrating ~4.0 Å rather than 2.5 Å resolution quality, and by a fairly poor omit ED for the nucleic acids (Supplementary Figs 7a and 8a). Careful analysis revealed that the data were affected by strong 'mis-scaling' between the low-resolution (~20–7 Å) and high-resolution data. Because this problem could not be resolved through a procedure

of the bulk solvent correction^{38,43} that uses the uniform solvent approach, we presumed that in the EC crystals the solvent might be non-uniform, thereby providing systematic, but distinct, contributions to the intensities in different resolution ranges. To test this hypothesis, we devised an original procedure in which 'zonal scaling' was applied to the diffraction data before the refinement and modelling of the nucleic acids in the EC using the unbiased (right after the molecular replacement) RNAP coordinates to eliminate or minimize the potential resolution-dependent systematic errors in the experimental data that might arise, in particular, from the non-uniformity of the solvent in the crystals. The EC data were divided into ~180 resolution shells (to allow for ~2,000 reflections per shell), and the individual scale factors between the observed and model structure factor amplitudes were determined and applied to the experimental data for each shell. The zonal scaling protocol is not entirely new and was previously implemented in the program PRISM⁴⁴ that used skeletonization of the ED as a technique of phase improvement to provide a better match between the observed and calculated structure factor amplitudes and was used successfully by us in the determination of the T4 endonuclease V/DNA complex structure⁴⁵. Moreover, another commonly accepted scaling procedure, 'local' scaling, although different in details, uses the same idea of the determination of the individual scale factor for a particular reflection through scaling applied to only a limited subset of the reflections adjacent to this reflection in a reciprocal space⁴⁶. This protocol would therefore apparently introduce a similar, resolution-dependent correction to the structure factor amplitudes as zonal scaling. Consistently, application of local scaling to scale the native and heavy-atom derivative data was reported to make a significant improvement in the signal-to-noise ratio in the difference Patterson maps⁴⁷, while this option is incorporated in the major programs using direct methods to localize the heavy-atom sites in the derivatives of protein crystals, where it is used to obtain a better estimate for the normalized structure factor amplitudes^{48,49}. The novelty of our approach is that we used the zonal scaling procedure towards a refinement of the atomic model.

Zonal scaling, in combination with the bulk solvent correction, led to a marked decrease in the initial R -factors (particularly at low resolution) and a smooth rigid-body refinement using the CNS program³⁸ that converged to an R -factor/ R_{free} of ~39% after B -factor refinement (Supplementary Table 2). Identification of merohedral twinning coupled with the non-crystallographic symmetry similar to that observed previously^{37,38,50} (Supplementary Table 3) additionally improved the R -factors to ~33% at 2.5 Å resolution just through the rigid-body refinement and provided a high-quality ED for both the existing protein model and, most importantly, for the nucleic acid chains missing from the model (Supplementary Figs 7b and 8b). This ED allowed us to replace the β' -subunit dispensable region (β' 165–449) mistraced in the RNAP holoenzyme with the proper model⁵¹, to model all the nucleic acid in the EC unambiguously and to obtain a final high-quality model on refinement³⁸ (alternating with manual model building³⁹) of the whole complex against the 2.5-Å resolution data (R -factor/ R_{free} = 0.238/0.267; Supplementary Table 2). The final model includes the RNAP core enzyme (α_2 , β , β' and ω subunits, molecular mass ~400 kDa), the 13 bp of the dwDNA, 9 bp of the RNA/DNA hybrid, and seven nucleotides of the displaced RNA transcript (Fig. 1 and Supplementary Fig. 1).

All available corrections to the experimental data and the atomic parameters used during refinement of the macromolecular structures are aimed at decreasing the crystallographic R -factor and subsequently increasing the quality of the ED map. Among these, those variable parameters against which the major minimization is performed (coordinates, B -factors) might be overfitted during the course of the refinement, resulting in a so-called 'model bias'. There are at least two solid criteria that allow us to detect and avoid a model bias: R_{free} , which would begin increasing, and the omit ED maps, which would become worse on overfitting of the model. Zonal scaling demonstrated an excellent match to both of these criteria. Indeed, this correction applied only once to a partial unrefined model resulted in a decisive decrease in the initial R_{free} value by ~6%, while the refinement against the corrected experimental data yielded an additional overall decrease in R_{free} by ~18% (note that the refinement against the original data set did not converge at all; Supplementary Table 2). In contrast, application of zonal scaling to the EC diffraction data permitted a marked improvement in the omit nucleic acid ED (Supplementary Fig. 8b) thereby providing additional evidence that the zonal correction does not introduce model bias but significantly improves signal-to-noise ratio. It is also worth noting that, although zonal scaling and bulk solvent corrections are both designed to provide a better match between the low-resolution and high-resolution data, their contributions to the structure factors are largely distinct, suggesting that these corrections may account for somewhat different properties of the crystals and are therefore complementary, rather than alternative, techniques. Indeed, application of the bulk solvent correction after zonal scaling resulted in an additional improvement in R_{free} (Supplementary Table 2).

Our preliminary data indicate that zonal scaling would probably aid in the refinement of many crystal structures and in some cases (like that of the EC structure) may emerge as a key technique for structure determination. The detailed analysis of this technique is now underway and the results on the application of zonal scaling to the refinement of the various crystal structures will be published elsewhere.

36. Kashkina, E. *et al.* Elongation complexes of *Thermus thermophilus* RNA polymerase that possess distinct translocation conformations. *Nucleic Acids Res.* **34**, 4036–4045 (2006).
37. Vassylyeva, M. N. *et al.* Purification, crystallization and initial crystallographic analysis of RNA polymerase holoenzyme from *Thermus thermophilus*. *Acta Crystallogr. D Biol. Crystallogr.* **58**, 1497–1500 (2002).
38. Brunger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.* **54**, 905–921 (1998).
39. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
40. Merrit, E. A. & Bacon, D. J. Raster3D: photorealistic molecular graphics. *Methods Enzymol.* **277**, 505–524 (1997).
41. Kraulis, P. J. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Cryst.* **24**, 946–950 (1991).
42. Esnouf, R. M. Further additions to MolScript version 1.4, including reading and contouring of electron-density maps. *Acta Crystallogr. D* **55**, 938–940 (1999).
43. Afonine, P. V., Grosse-Kunstleve, R. W. & Adams, P. D. A robust bulk-solvent correction and anisotropic scaling procedure. *Acta Crystallogr. D Biol. Crystallogr.* **61**, 850–855 (2005).
44. Baker, D., Bystroff, C., Fletterick, R. J. & Agard, D. A. PRISM: topologically constrained phased refinement for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **49**, 429–439 (1993).
45. Vassylyev, D. G. *et al.* Atomic model of a pyrimidine dimer excision repair enzyme complexed with a DNA substrate: structural basis for damaged DNA recognition. *Cell* **83**, 773–782 (1995).
46. Matthews, B. W. & Czerwinski, E. W. Local scaling: a method to reduce systematic errors in isomorphous replacement and anomalous scattering measurements. *Acta Crystallogr. A* **31**, 480–497 (1975).
47. Wang, J. H. *et al.* Structure of a functional fragment of VCAM-1 refined at 1.9 Å resolution. *Acta Crystallogr. D Biol. Crystallogr.* **52**, 369–379 (1996).
48. Schneider, T. R. & Sheldrick, G. M. Substructure solution with SHELXD. *Acta Crystallogr. D Biol. Crystallogr.* **58**, 1772–1779 (2002).
49. Weeks, C. M. & Miller, R. Optimizing Shake-and-Bake for proteins. *Acta Crystallogr. D Biol. Crystallogr.* **55**, 492–500 (1999).
50. Yeates, T. D. Detecting and overcoming crystal twinning. *Methods Enzymol.* **276**, 344–358 (1997).
51. Chlenov, M. *et al.* Structure and function of lineage-specific sequence insertions in the bacterial RNA polymerase β' subunit. *J. Mol. Biol.* **353**, 138–154 (2005).

Structural basis for substrate loading in bacterial RNA polymerase

Dmitry G. Vassylyev¹, Marina N. Vassylyeva¹, Jinwei Zhang², Murali Palangat³, Irina Artsimovitch⁵
& Robert Landick^{3,4}

The mechanism of substrate loading in multisubunit RNA polymerase is crucial for understanding the general principles of transcription yet remains hotly debated. Here we report the 3.0-Å resolution structures of the *Thermus thermophilus* elongation complex (EC) with a non-hydrolysable substrate analogue, adenosine-5'-[(α,β)-methyleno]-triphosphate (AMPcPP), and with AMPcPP plus the inhibitor streptolydigin. In the EC/AMPcPP structure, the substrate binds to the active ('insertion') site closed through refolding of the trigger loop (TL) into two α -helices. In contrast, the EC/AMPcPP/streptolydigin structure reveals an inactive ('preinsertion') substrate configuration stabilized by streptolydigin-induced displacement of the TL. Our structural and biochemical data suggest that refolding of the TL is vital for catalysis and have three main implications. First, despite differences in the details, the two-step preinsertion/insertion mechanism of substrate loading may be universal for all RNA polymerases. Second, freezing of the preinsertion state is an attractive target for the design of novel antibiotics. Last, the TL emerges as a prominent target whose refolding can be modulated by regulatory factors.

The mechanism of substrate selection and loading into the active site is of central importance in the studies of the multisubunit RNA polymerases (RNAPs), which mediate gene expression in all cellular organisms and in which substrate loading is tightly coupled to the overall processivity, fidelity and regulation of transcription. Not only will understanding of this mechanism illuminate the basic principles of transcription, but it will also have important medical applications by facilitating the design of novel antibiotics targeting the nucleotide addition cycle (NAC) of bacterial RNAP.

Structural snapshots of the substrate-bound transcription intermediates obtained by X-ray crystallography have provided a clear view of the NAC in single-subunit RNAPs, suggesting a two-step substrate loading mechanism^{1,2}. In the T7 EC, the incoming substrate NTP first binds to the inactive ('preinsertion') conformation. In the preinsertion state the substrate is base-paired with the acceptor template base while the catalytic O/O' helices anchoring the NTP phosphates are in the open configuration, thereby positioning the substrate several ångströms from the catalytic centre. Second, the EC isomerizes into the catalytically active ('insertion') state by closure of the active site after rotation of the catalytic O/O' helices and the concerted movement of the substrate into the catalytic centre.

In contrast, the pathway of substrate loading into the active site of multisubunit RNAPs remains a matter of debate³. Although an analogous two-step mechanism was proposed for bacterial enzymes^{1,4}, one group working on a structural analysis of eukaryotic RNAPs reported findings of a templated NTP in the preinsertion site⁵, whereas another emphasized a template-independent site of entry⁶ (overlapping, but distinct from, a previously proposed E-site whose existence was inferred from an analysis of RNAP active-site exonuclease activity, in the location of the nucleotide base⁷) while ruling out the existence of the preinsertion state. Moreover, 'multisubstrate' models based on kinetic analyses posit the existence of at least two other sites to which substrate can bind simultaneously with the insertion site^{8,9}. These apparently irreconcilable models of substrate

selection and loading underscore the need for additional investigation including high-resolution structural analysis.

We have determined the crystal structures of the *T. thermophilus* EC (ttEC¹⁰) with the non-hydrolysable substrate analogue AMPcPP, and with AMPcPP and antibiotic streptolydigin (Stl) at 3.0 Å resolution (Fig. 1, Supplementary Fig. 1 and Supplementary Table 1). Our results argue that NTP loading occurs in two steps: first, substrate binding to an open, inactive ('preinsertion') state, and second, refolding of the TL into two trigger helices (TH) that stabilize the closed, catalytically active ('insertion') intermediate (Fig. 1). Stl induces displacement of the TH, thereby freezing the EC in the inactive state (Fig. 1).

Insertion site: closing by folding of the TL

In the ttEC/AMPcPP complex (Fig. 1a), formation of a closed substrate-bound intermediate occurs on folding of the TL (β' 1236–1254) into two α -helical segments that extend the α -helices flanking the TL to form two long anti-parallel helices (TH, β' 1221–1265); together with the adjacent bridge helix (BH, β' 1066–1103), this creates a three-helical bundle (Fig. 2a and Supplementary Fig. 2a). TL refolding reduces the dimensions of the secondary channel¹¹ at the entrance to the active site from $\sim 15 \text{ Å} \times 22 \text{ Å}$ in the substrate-free EC to $\sim 11 \text{ Å} \times 11 \text{ Å}$ in the substrate-bound EC (Supplementary Fig. 2b, c). This constriction probably hinders substrate dissociation and competition but would still permit the entry of regulators that are thought to directly remodel the RNAP catalytic centre on binding through the secondary channel¹². Closure of the active site, although not accompanied by the TL \rightarrow TH transition, was also proposed to occur in eukaryotic EC¹³.

In the substrate-free EC, where TL is disordered¹⁰, certain structural elements may adopt a conformation that is unfavourable for TH formation. Superposition of the ttEC and ttEC/AMPcPP structures (root mean squared deviation 0.37 Å over the 500 C α atoms) revealed local alterations near the active site apparently triggered by substrate binding. First, the rearrangement of the downstream DNA (dwDNA) is accompanied by a repositioning of fork loop 2 (fork2, β 413–451)

¹Department of Biochemistry and Molecular Genetics, University of Alabama at Birmingham, Schools of Medicine and Dentistry, 402B Kaul Genetics Building, 720 20th Street South, Birmingham, Alabama 35294, USA. ²Department of Biomolecular Chemistry, ³Department of Biochemistry and ⁴Department of Bacteriology, University of Wisconsin, Madison, Wisconsin 53706, USA. ⁵Department of Microbiology, The Ohio State University, 484 West 12th Avenue, Columbus, Ohio 43210, USA.

(Supplementary Fig. 3b) that is likely functionally significant: the fork 2 amino-terminal hairpin stacks on the ($i+2$) base pair, whereas its carboxy-terminal portion contributes to the formation of the closed active site, thereby bridging the dwDNA and substrate in the EC¹⁰. Second, a ~ 1.5 -Å shift of the central portion of the BH (β' 1077–1095) positions the BH between the TH, so that the BH side chains fit into the grooves of these helices. This BH displacement (which differs from the BH bending/unfolding proposed to control translocation and NTP loading^{14–16}) seems crucial for the TL \rightarrow TH transition: it supports an optimal network of contacts within the three-helical bundle while eliminating steric clash that would occur between the TH and BH positioned as in the substrate-free EC (Supplementary Fig. 3b, c; Phe β' 1241 from TH and Thr β' 1084 from BH would produce the major clash; an alignment of all the residues discussed in this paper with *Escherichia coli* and yeast RNAPs is presented in Supplementary Table 2).

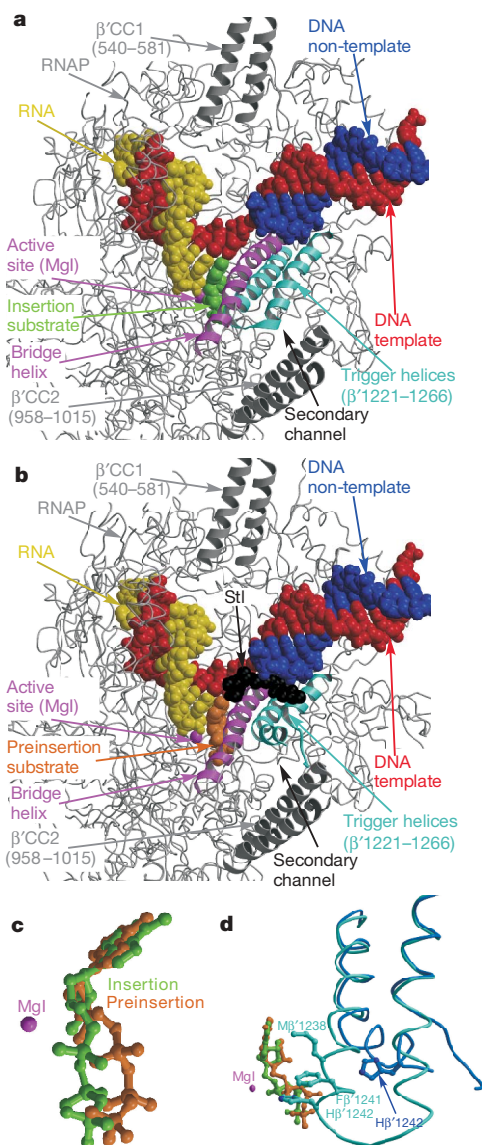


Figure 1 | Structures of the substrate complexes. The same colour scheme is used in all figures. The DNA template, non-template and RNA strands are in red, blue and yellow, respectively. The BH, the TH and the rest of the RNAP molecule are in magenta, cyan and grey, respectively. The insertion and preinsertion NTP analogues and Stl are designated by green, orange and black, respectively. The catalytic Mg^{2+} ions (MgI and MgII) are shown as magenta spheres. **a, b**, Overall views of the ttEC/AMPCPP (**a**) and EC/AMPCPP/Stl (**b**) complexes. CC, coiled coil. **c, d**, Superposition of the NTPs (**c, d**) and TH (**d**) in the insertion (green, NTP; cyan, TH) and preinsertion (orange, NTP; blue, TH) complexes.

Insertion site: substrate binding

In the ttEC/AMPCPP complex, AMPCPP forms a Watson–Crick base pair with the acceptor template sandwiched between the 3' end of the RNA/DNA hybrid on one side and the protein residues from the TH and BH on the other. Met β' 1238 (TH) and Thr β' 1088 (BH), which stack directly on the substrate and template bases, respectively, seem essential for the positioning and selection of the substrate (Fig. 2); the corresponding RNAPII residues Leu 1081 and Thr 831 seem to have similar functions in the yeast EC¹³. In agreement with modelling and biochemical studies^{1,4,15}, Asn β' 737 forms hydrogen bonds with both O3' and O2' of the substrate ribose, thereby permitting discrimination against the non-cognate dNTPs lacking either or both of these atoms (Fig. 2b, c). This is in contrast with the eukaryotic EC structures in which the Asn β' 737 counterpart interacts only with the O3' atom¹³. However, in agreement with the eukaryotic structures, Arg β' 704 forms hydrogen bonds bridging the O2' atom of the substrate and the 3' RNA nucleotide (Fig. 2c and Supplementary Fig. 4), which is consistent with its proposed role in discriminating between rNTP and dNTP.

The substrate phosphates reside in the active site and adopt a folded conformation that permits the concerted coordination of both the high-affinity (MgI) and low-affinity (MgII) catalytic Mg^{2+} ions, which are located ~ 3.9 Å apart (as was also observed in the high-resolution 'insertion' structures of T7 RNAP and DNA polymerase^{2,17}). All three phosphates coordinate MgII, and the α -phosphate additionally binds MgI. As expected, the catalytic Asp triad (β' 739, β' 741 and β' 743) provides three ligands to MgI (Fig. 2d and Supplementary Figs 4a and 5). However, in contradiction of the expectations that MgII would be directly coordinated by two β' triad Asp residues and by the invariant Asp β 686 side chain^{6,18}, only one triad residue, Asp β' 739, bridges both Mg^{2+} ions (Fig. 2d and Supplementary Figs 4a and 5). A similar coordination pattern was observed in DNA polymerases^{2,17}.

The electron density (ED) reveals a water molecule near MgII (Supplementary Fig. 6) whose position is fixed by the substrate phosphates

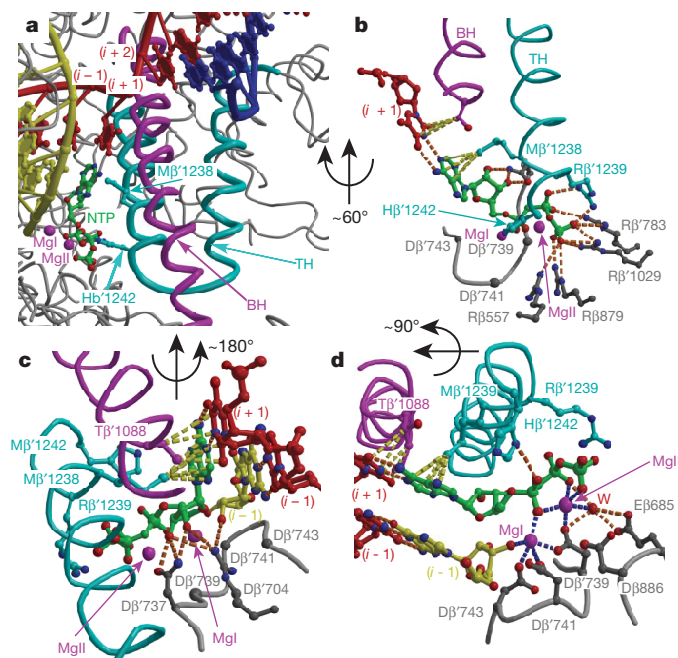


Figure 2 | The insertion complex. **a**, Close-up view through the main channel of the folded TH closing the RNAP active site. **b–d**, Three views showing a three-dimensional pattern of protein direct (**b**) and Mg^{2+} -mediated (**d**) interactions with the AMPcPP phosphates, discriminating contacts with the ribose (**c**) and stacking interactions with the base (**b–d**) of AMPcPP. The rotations necessary to obtain the views in **b** and **c** relative to **a** and of **b** relative to **d** are indicated. Hydrogen bonds, van der Waals interactions and Mg^{2+} coordination bonds are shown by orange, yellow and blue dashed lines, respectively.

and several protein residues including invariant Glu β 685 (W in Fig. 2d and Supplementary Figs 4a and 5). In the final structure, two ligands for the metal ions are missing (one for each Mg^{2+} ion); however, their positions can be inferred because all other bonds agree well with the octahedral Mg^{2+} coordination. Modelling shows that both missing bonds may be donated by a single water molecule located within the interacting distance of Asp β 686 (pW in Supplementary Fig. 7), which may have a catalytic function. The direct and indirect (water-mediated) coordination of the catalytic metal ions by the β' triad and the β acidic pair, respectively, is consistent with the observations that replacements of any of the β' Asp residues^{19–21} confer a much more deleterious effect than those of the β residues^{7,22–25}.

The conformation of the substrate phosphates is stabilized through 11 hydrogen bonds with six basic residues, among which four were predicted from the yeast EC structure⁵ and modelling of the bacterial substrate complex¹; the other two, Arg β' 1239 and His β' 1242, belong to the TH (Fig. 2b and Supplementary Fig. 4a). Together with two additional TH residues, Met β' 1238 and Phe β' 1241, these residues form a complex network of interactions with the substrate that may determine the optimal shape of the closed active-site cavity and favour the proper loading of cognate NTPs (see below) while excluding non-cognate NTPs.

Taken together, the highly specific mode of substrate binding, the arrangement and coordination of the catalytic Mg^{2+} ions, the stable closure of the active site by the TH, and the apparent similarity to functionally homologous systems^{2,17} strongly argue that the ttEC/AMPcPP structure corresponds to the catalytically active insertion transcription intermediate.

Preinsertion state: Stl-induced displacement of the TH

Our structural and functional analyses of the mechanism of Stl²⁶, which inhibits bacterial RNAP during both initiation and elongation, led us to conclude that Stl does not compete with the substrate but rather stabilizes the inactive, substrate-bound preinsertion state, presumably by interfering with closure of the active site (but see ref. 27). The ttEC/AMPcPP/Stl structure confirms this prediction: whereas both Stl and AMPcPP were bound to the ttEC, the TH are only partly folded and are displaced from their position in the ttEC/AMPcPP complex by as much as 15 Å, leaving the active site open (Fig. 3a).

Stl binds in the dwDNA channel along the BH, ~ 20 Å and ~ 10 Å from the β' Asp triad and substrate base, respectively (Fig. 3a and Supplementary Fig. 8). The Stl contacts to the ttEC are different from those in the RNAP/Stl complex²⁶ by a few polar and multiple van der Waals interactions that Stl forms along its entire length with the dwDNA (Supplementary Fig. 8b, c). The effect of Stl on TH displacement is probably indirect and consists of two major components. First, binding of the streptolol ring in the pocket formed by the BH fragment (β' 1082–1086) and fork2 and its van der Waals contacts with the acceptor base would hinder their movements after substrate binding in the insertion complex; in particular, the BH in the Stl-induced conformation would probably clash with the closed TH, as described above for a substrate-free EC. Second, on the opposite side of the Stl molecule, the acetamide moiety stacks on Ile β' 1260 and Pro β' 1257, slightly displacing the latter (Supplementary Fig. 8b, c); these interactions might disturb the network on the interhelical TH contacts required for the proper folding of this TL segment. As a result, the TL C terminus (β' 1251–1255) is in an extended rather than an α -helical conformation, whereas the TH tip (β' 1244–1250) seems disordered (Figs 1d and 3a). The N-terminal half of the TL retains the α -helical configuration; however, the helix is sharply bent, allowing it to pack in the dwDNA channel near Stl (and far from the substrate), and makes extensive hydrophobic interactions with both the BH and Stl.

Preinsertion state: substrate binding

Apart from TL displacement from the active position, Stl binding does not affect any structural elements essential for substrate binding in the active site. Consistently, our structural data show that substrate

may still bind with high affinity and specificity in the open conformation. However, the open intermediate seems catalytically inactive as a result of the displacement of the substrate phosphates (a $\sim 35^\circ$ rotation occurring approximately about the C4'–C5' bond resulting in a ~ 2.5 Å overall shift; Fig. 1c). Although still in the folded configuration that is capable of anchoring the MgII, the phosphates reside too far from the Mgl (~ 5.5 Å to the α -phosphate) to permit the catalytic reaction (Fig. 2b–d and Supplementary Figs 4b and 9). Given that the base-specific and ribose-specific interactions are nearly identical to those in the insertion complex (see below), the ttEC/AMPcPP/Stl structure probably represents a structural snapshot of a template-dependent inactive substrate complex. This complex may be functionally similar to the T7 preinsertion complex, even though the difference in preinsertion and insertion substrate positions is less marked than was observed in the T7 EC.

Whereas the absence of the folded TH probably accounts for the inactive orientation of the triphosphate, its binding is mediated by 15 hydrogen bonds with four basic protein side chains (Fig. 3b) and should be at least as stable as in the insertion complex. Among these, Arg β 557 donates only one hydrogen bond to the substrate in the insertion complex vs eight hydrogen bonds bridging the α - and γ -phosphates in the preinsertion complex, thereby probably compensating for the lack of the polar interactions with the TH in the preinsertion state (Fig. 3b and Supplementary Fig. 4). Four basic residues form a rim around the NTP phosphates in both the preinsertion and insertion states (Arg β 557, Arg β 879, Arg β' 783 and Arg β' 1029; Figs 2b and 3b, and Supplementary Fig. 4)^{1,5}. The rim is stabilized through stacking interactions between the Arg side chains that largely determine and restrain their orientations. Replacement of one of these residues may increase their mobility and induce rearrangement of the remaining basic residues in the rim that would allow them to maintain the same number of contacts with NTP as in the wild-type enzyme. Indeed, the substitution of Arg β 557 (β R678 in *E. coli*) for Cys does not have large effects on

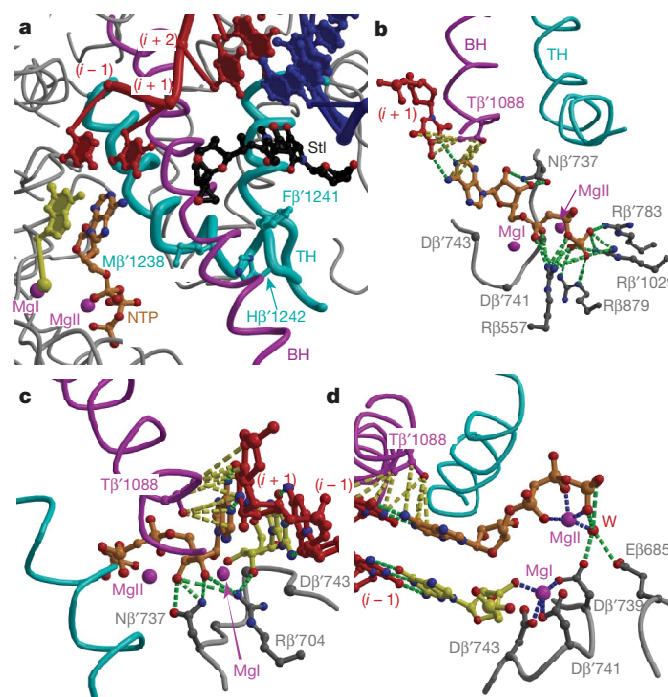


Figure 3 | The preinsertion complex. **a**, The TH in the Stl-bound EC. **b–d**, Three views (in similar orientations to those in Fig. 2b–d) showing a three-dimensional pattern of protein direct (**b**) and Mg^{2+} -mediated (**d**) interactions with the AMPcPP phosphates, discriminating contacts with the ribose (**b**) and stacking interactions with the base (**b–d**) of AMPcPP. Hydrogen bonds, van der Waals interactions and Mg^{2+} coordination bonds are shown by green, yellow and blue dashed lines, respectively.

the substrate affinity or the catalytic rate²⁵. Whereas in the insertion state all three NTP phosphates have protein ligands, in the preinsertion state the β -phosphate lacks interactions with the protein; this suggests that the triphosphate adopts a somewhat relaxed conformation that resists loading into the catalytic position. The basic rim may therefore serve as a 'gate' that prevents catalysis in the open EC conformation. Tyr 639, which sterically blocks the active site in the preinsertion state, may have a similar function in T7 RNAP¹. Isomerization into the insertion state would require the gate to 'open', probably in response to TH folding that might alter the NTP interactions to facilitate its release from the preinsertion contacts and at the same time stabilize the insertion configuration. His β '1242 might be important for this transition, because in the TH configuration it clashes most severely with the preinsertion NTP phosphates; however, in the model the clash also occurs with the TH backbone around position β '1239, suggesting that TH folding may be the key for the preinsertion/insertion transition.

Despite the lack of contacts with the TH in the open active site, the base and sugar substrate moieties occupy positions similar to those in the insertion state (root mean squared deviation ~ 0.6 Å); thus, both Watson–Crick base pairing with the template and ribose-discriminating contacts with Asn β '737 (Fig. 3b–d and Supplementary Fig. 4b) are already established in the preinsertion state. Interestingly, Thr β '1088 (BH), which in the insertion complex stacks only on the acceptor template base (Fig. 2), forms additional stacking interactions with the NTP base in the preinsertion state that seem to compensate for the absence of Met β '1238 (TH). The BH and, in particular, Thr β '1088 may therefore serve as an important fidelity determinants in both the preinsertion and insertion states. Consistently, substitution of Thr β '1088 for Val, which would enhance van der Waals interactions with the NTP/DNA template base pair, also enhanced selectivity for rNTP over dNTP⁴. The effect of Asp β 554 mutations on transcription fidelity²⁵ may be also explained by the repositioning of the BH resulting from these substitutions: in both complexes Asp β 554 resides far from the NTP but in close proximity to the BH.

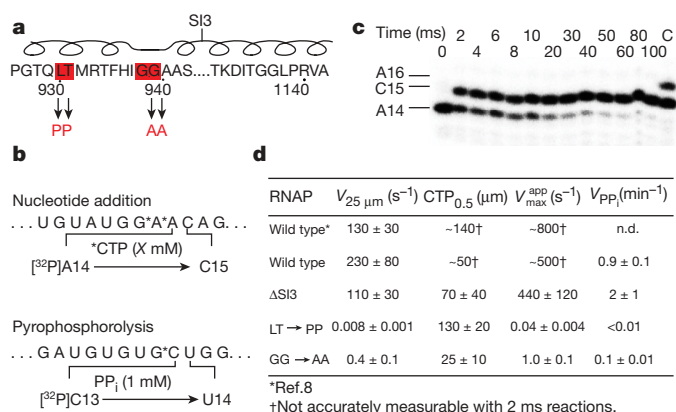


Figure 4 | Effect of TL alterations on nucleotide addition rate. **a**, Location of TL alterations in the *E. coli* RNAP. The diagram above the sequence indicates the position of the TH. SI3 denotes the location of SI3 in wild-type *E. coli* RNAP. *E. coli* LT930,1 and GG938,9 correspond to *T. thermophilus* LT1236,7 and GG1244,5, respectively. **b**, Sequence of the transcription scaffolds surrounding the sites used to assay nucleotide addition and pyrophosphorolysis (see Methods). C15 corresponds to C25 in ref. 8. 'X mM CTP' represents the range of CTP concentrations used in the assay. **c**, Electrophoretic separation of reaction products from a quench-flow assay of wild-type RNAP at 25 μ M CTP (see Methods). **d**, Kinetic parameters measured for wild-type, Δ SI3, PP and AA RNAPs. n.d., not determined. Values estimated from Fig. 1C in ref. 8 are shown for comparison. CTP_{0.5}, the concentration of CTP giving half-maximal velocity, is reported rather than K_{NTP} because equilibrium binding may not be a valid assumption. The rate of pyrophosphorolysis (V_{PPi}) was determined at 1 mM PP_i. Errors are s.e.m. and were calculated as described in Methods.

TL folding is essential for catalysis

The TH are apparently less flexible than the unfolded TL, because their conformation is additionally stabilized through intrahelical and interhelical interactions, as well as through contacts with the adjacent protein segments. However, the TH may be unstable in the absence of the latter contacts, suggesting that the substrate not only anchors the TL, as implied by the eukaryotic TL model¹³, but may also induce changes that create a network of the protein–protein contacts along the entire TH length to stabilize the flexible portion of the TL. Indeed, even subtle alterations would disrupt these crucial interactions and could alter folding/positioning of the TH even in the presence of the substrate (for example, as observed in the ttEC/AMPcPP/Stl structure).

To test the hypothesis that folding of the TL, rather than its movement in the unfolded state¹³, is a crucial component of the substrate-loading mechanism in bacterial RNAPs, we created substitutions within the TH that do not involve the substrate-interacting residues and should not affect seriously the movement and conformation of the unfolded TL, but would instead have a specific effect on the TL \rightarrow TH transition (Leu–Thr \rightarrow Pro–Pro and Gly–Gly \rightarrow Ala–Ala; Fig. 4a). Replacement of Leu–Thr with Pro–Pro (LT \rightarrow PP) should completely block TL folding, whereas replacement of Gly–Gly with Ala–Ala (GG \rightarrow AA) should restrict flexibility of the turn between the folded TH and perturb their packing and folding. We purified mutant *E. coli* RNAP enzymes containing these substitutions but lacking SI3, a 188-residue insertion that is absent from the *T. thermophilus* TL. Deletion of SI3 has only modest effects on nucleotide addition²⁸. We tested the effect of the PP and AA substitutions on the rate of CMP addition in TECs directly reconstituted on a DNA/RNA scaffold (see Methods), using a sequence previously found to support rapid nucleotide addition (Fig. 4b)⁸. With wild-type *E. coli* RNAP, we observed rates at saturating, half-maximal and 25 μ M CTP that were comparable to those reported previously⁸ but too fast for kinetic analysis (Fig. 4c and Supplementary Fig. 10). Deletion of SI3 slowed nucleotide addition to a measurable but still rapid range. However, the PP and AA substitutions slowed CMP addition markedly, with the PP mutant $\sim 10^4$ slower than Δ SI3 RNAP, even at saturating CTP, and the AA mutant ~ 400 times slower (Fig. 4d). However, little difference was seen in the concentration of CTP required for half-maximal velocity, suggesting the principal defect was in catalysis rather than in substrate binding. Large defects in rates of pyrophosphorolysis (Fig. 4d) are consistent with this interpretation. The large effects on catalytic rate with minimal effects on substrate affinity of these substitutions argue strongly for the existence of the preinsertion state deduced from our structural analysis.

Substrate-induced opening of the dwDNA clamps

Superposition of the ttEC and ttEC/AMPcPP structures by the dwDNA atoms (root mean squared deviation ~ 0.2 Å) revealed a substantial opening of the dwDNA clamps that results in the ~ 4 Å displacement of the β -pincer (whose orientation in the substrate complex corresponds well to that observed in the 'open' holoenzyme structure²⁹), with a concomitant loss of several polar and van der Waals interactions with the dwDNA in the substrate-bound complex (Supplementary Fig. 11). Because no such opening was observed in the ttEC/AMPcPP/Stl complex, we conclude that these alterations may arise from changes near the active site that occur on isomerization to the closed, insertion state. Indeed, in both the substrate-free and substrate-bound complexes, the N-terminal segment of fork 2 forms an extensive hydrophobic interface with the β -pincer, suggesting that its repositioning is a consequence of substrate-dependent fork 2 movement. Thus, along with TH formation, opening of the dwDNA clamps seems to be a hallmark of the closed insertion complex. The insertion complex remains stable as a result of substrate binding despite the increased dwDNA mobility; however, it may become prone to translocation after addition of the nucleotide and formation of the pre-translocation complex (see below).

Nucleotide addition cycle

We propose the following pathway for the NAC in the multisubunit enzymes (Fig. 5). Before NTP loading, the EC exists in equilibrium between the pre-translocated and post-translocated states, whose inter-conversion is driven by thermal motions. In the pre-translocated EC, where the 3' RNA nucleotide occupies the position taken by the substrate in the post-translocated EC, the TL→TH transition might have a function in the exo-nucleolytic activity of RNAP⁷ (by promoting closure of the active site) or in facilitating translocation (by promoting the opening of the dwDNA claws, which probably accompanies TH formation). Substrate loading occurs in two steps. First, the NTP binds to the open EC to form an inactive, preinsertion intermediate, working as a 'ratchet' that stabilizes the post-translocated EC³⁰ and providing the first sieve for substrate selection. Second, on folding of the TL into the TH, the NTP is repositioned to form the catalytically competent insertion complex, in which the active site is closed. The inherent instability of the TH, the opening of the dwDNA claws, and the potential competition between the preinsertion and insertion contacts for the substrate triphosphate suggest that the insertion complex may be a high-energy state that is more sensitive to the shape or chemical properties of the incoming NTP than the preinsertion EC, thereby providing the second, finer sieve for substrate selection. The NAC culminates with the catalytic reaction, resulting in transcript extension and pyrophosphate release. The loss of the phosphate moieties then destabilizes the insertion complex and triggers relaxation to the ground state, which may allow transition from the closed, pre-translocated state to an open, presumably post-translocated, state² characterized by the unfolded TL and closed β -pincer, as observed in the substrate-free eukaryotic^{5,13} and bacterial ECs¹⁰. However, this is unlikely to create a 'power-stroke'-like mechanism² because the structural alterations do not seem to form a 'ratchet' that would stabilize the post-translocated state irreversibly¹⁰, and physical measurements³⁰ also suggest that translocation is best explained by an NTP-biased, brownian-ratchet model.

Comparison with T7 and eukaryotic RNAPs

In spite of their drastically different architectures, multisubunit bacterial and single-subunit phage ECs use structurally analogous strategies to handle the nucleic acids¹⁰. The substrate-bound tTEC

structures revealed additional similarities between these systems. First, in both enzyme classes substrate loading occurs in two steps, in which the inactive, preinsertion state probably provides quality control for the initial substrate before its delivery to the insertion state. Second, both systems use preinsertion 'gates' to prevent premature substrate alignment in the catalytic centre. Third, both types of insertion complex might represent high-energy states used for the final selection of the cognate substrate. Last, the structural elements that control the transition from open to closed configurations seem strikingly similar in both conformation (two anti-parallel α -helices called TH and O/O'-helices in the tTEC and T7 EC, respectively) and mode of substrate binding (key hydrophobic and basic residues located in the N-terminal helices bind to the substrate aligned roughly parallel to the helical axes; Supplementary Fig. 12). However, these catalytic helices have different functions in substrate loading. In the T7 EC, the folded O/O' helices anchor the substrate in the preinsertion state, positioning the substrate and acceptor DNA template far from the active site. In contrast, in the preinsertion bacterial EC the TH are probably unfolded and bind the substrate only after their folding in the insertion complex. Moreover, whereas TH function is apparently limited to the proper positioning of the NTP in the active site, the O/O' helices also contain the 'gate' residue Tyr 639 and an important fidelity determinant, Phe 644 (ref. 31), which is involved in the separation of the dwDNA strand.

Several medium-resolution yeast EC structures bound to the NTP substrates and their analogues have been reported^{5,6,13}. As expected on the basis of the significant structural and functional similarity among all multisubunit RNAPs, our analysis of the bacterial EC has revealed numerous similarities with the yeast structures, including that with a proposed preinsertion mechanism of substrate loading⁵. Nonetheless, several key aspects of the bacterial substrate complexes differ from the eukaryotic closed EC/NTP complex¹³. The most significant differences include the dwDNA configuration and contacts, the TL conformations in the closed complexes, and the binding modes/sites for the NTP substrate. We think that these discrepancies (analysed in the Supplementary Discussion) are due largely to the differences in the crystallization conditions and nucleic-acid scaffolds, to difficulties in the interpretation of the lower-resolution structural data available for the eukaryotic complexes, or both.

A detailed comparison of the relevant bacterial and eukaryotic structures (Supplementary Discussion and Supplementary Figs 13–15) argues for the existence of a catalytically inactive preinsertion state that may have key functions in all RNAPs. First, the preinsertion substrate may act as a ratchet to stabilize the post-translocated state. Second, the substrate is able to be initially surveyed in (and rejected from) the open RNAP configuration. Third, the preinsertion state prevents premature catalysis in the open state. Last, stabilization of NTP in the preinsertion state might be used as a response to the lesions present in the dwDNA (see Supplementary Discussion). The alternative template-independent substrate entry site that was observed experimentally⁶ cannot have any of these functions. However, the entry site may still function not as an alternative but rather as an additional, third substrate intermediate en route to the preinsertion state²⁶.

Taken together, the structural analysis and the very high conservation of the active site residues in all kingdoms of life imply that the eukaryotic and bacterial RNAPs use the same mechanism of substrate loading and catalysis. The rigorous, detailed comparison between the molecular mechanisms of bacterial and eukaryotic RNAPs awaits the structural explanation of ECs assembled on identical, more natural scaffolds for both systems.

Concluding remarks

Our conclusion that the isolated TH conformation may require stabilization through interactions with both the substrates and numerous adjacent RNAP structural elements has several implications. First, all reactions catalysed by RNAP (nucleotide addition, exo-nucleolytic and endonucleolytic cleavages, and pyrophosphorolysis)

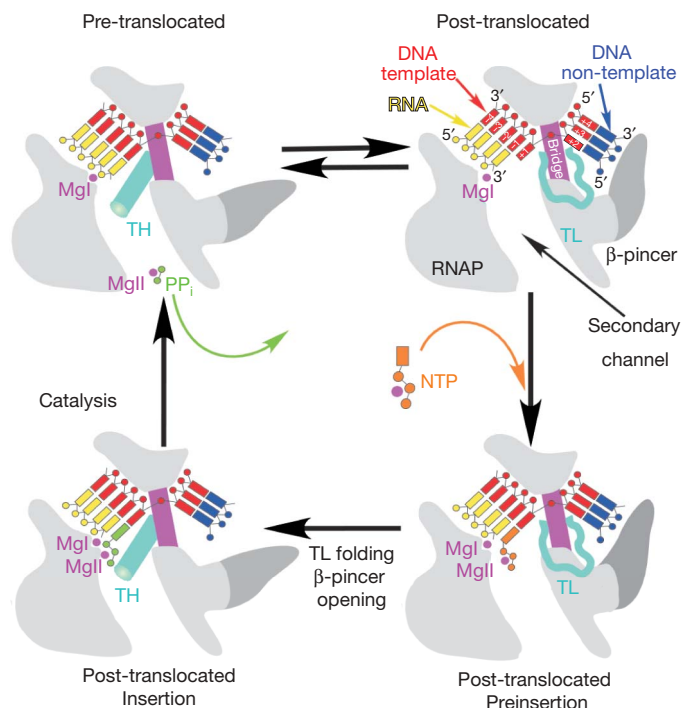


Figure 5 | Nucleotide addition cycle. The substrate loading pathway in bacterial RNAP.

probably require the TL→TH transition if they are to achieve the closed, catalytically active state. This implies that the extrinsic factors that markedly increase the RNAP intrinsic exonucleolytic (non-complementary NTP) and endonucleolytic (Gre factors) activities not only donate the missing ligands to the catalytic Mg^{2+} ions^{7,32,33} but may also interact with the TH (directly or indirectly) to stabilize the closed intermediate. Indeed, the eukaryotic counterpart of the Gre-factors, TFIIS, interacts with the TL⁵, and the TL→TH conversion may be required for the proper positioning of its 'catalytic' acidic residues in the RNAP catalytic centre. Second, changes in the geometry of the RNAP active site that are thought to accompany the formation of paused ECs probably involve an alternative TL/TH configuration stabilized in part through interactions of the frayed 3' RNA nucleotide with the residues in the E-site, which may be formed in part by TL/TH residues³⁴. Third, various auxiliary factors might activate long-distance allosteric signals modulating the TL↔TH switch, thereby conferring a range of regulatory effects. Thus, along with the catalytic residues and Mg^{2+} ions, the dynamic TL/TH structure emerges as a central regulatory element in the transcription complexes. Our observation that the antibiotic Stl blocks the closure of the active site and freezes the EC in the inactive state without even making direct polar interactions with the TL opens a range of opportunities for the design of novel antibiotics targeting the TL→TH transition in bacterial RNAPs.

METHODS SUMMARY

Crystallization of both complexes was performed under essentially the same conditions as for the ligand-free EC¹⁰. The structures were refined to a final R -factor/ R_{free} ratio of 0.225/0.257 and 0.234/0.266 for the tTEC/AMPcPP and tTEC/AMPcPP/Stl structures, respectively, at 3.0 Å resolution (Supplementary Table 1). Both nucleotide addition and pyrophosphorolysis reactions were performed with ECs reconstituted with *E. coli* RNAP and nucleic acid scaffolds as described in Supplementary Information. For nucleotide addition, kinetic parameters (V_{max} and $NTP_{0.5}$) were determined from reactions over a wide range of substrate concentrations without assumption of any particular reaction mechanism.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 28 January; accepted 11 May 2007.

Published online 20 June 2007.

1. Temiakov, D. *et al.* Structural basis for substrate selection by T7 RNA polymerase. *Cell* **116**, 381–391 (2004).
2. Yin, Y. W. & Steitz, T. A. The structural mechanism of translocation and helicase activity in T7 RNA polymerase. *Cell* **116**, 393–404 (2004).
3. Landick, R. NTP-entry routes in multi-subunit RNA polymerases. *Trends Biochem. Sci.* **30**, 651–654 (2005).
4. Svetlov, V., Vassilyev, D. G. & Artsimovitch, I. Discrimination against deoxyribonucleotide substrates by bacterial RNA polymerase. *J. Biol. Chem.* **279**, 38087–38090 (2004).
5. Kettenberger, H., Armache, K. J. & Cramer, P. Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIS. *Mol. Cell* **16**, 955–965 (2004).
6. Westover, K. D., Bushnell, D. A. & Kornberg, R. D. Structural basis of transcription: nucleotide selection by rotation in the RNA polymerase II active center. *Cell* **119**, 481–489 (2004).
7. Sosunov, V. *et al.* Unified two-metal mechanism of RNA synthesis and degradation by RNA polymerase. *EMBO J.* **22**, 2234–2244 (2003).
8. Foster, J. E., Holmes, S. F. & Erie, D. A. Allosteric binding of nucleoside triphosphates to RNA polymerase regulates transcription elongation. *Cell* **106**, 243–252 (2001).
9. Gong, X. Q., Zhang, C., Feig, M. & Burton, Z. F. Dynamic error correction and regulation of downstream bubble opening by human RNA polymerase II. *Mol. Cell* **18**, 461–470 (2005).
10. Vassilyev, D. G., Vassilyeva, M. N., Perederina, A., Tahirov, T. H. & Artsimovitch, I. Structural basis for transcription elongation by bacterial RNA polymerase. *Nature* doi:10.1038/nature05932 (this issue).
11. Zhang, G. *et al.* Crystal structure of *Thermus aquaticus* core RNA polymerase at 3.3 Å resolution. *Cell* **98**, 811–824 (1999).
12. Nickels, B. E. & Hochschild, A. Regulation of RNA polymerase through the secondary channel. *Cell* **118**, 281–284 (2004).
13. Wang, D., Bushnell, D. A., Westover, K. D., Kaplan, C. D. & Kornberg, R. D. Structural basis of transcription: role of the trigger loop in substrate specificity and catalysis. *Cell* **127**, 941–954 (2006).
14. Epshtein, V. *et al.* Swing-gate model of nucleotide entry into the RNA polymerase active center. *Mol. Cell* **10**, 623–634 (2002).
15. Gnatt, A. L., Cramer, P., Fu, J., Bushnell, D. A. & Kornberg, R. D. Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 Å resolution. *Science* **292**, 1876–1882 (2001).
16. Bar-Nahum, G. *et al.* A ratchet mechanism of transcription elongation and its control. *Cell* **120**, 183–193 (2005).
17. Doublier, S., Tabor, S., Long, A. M., Richardson, C. C. & Ellenberger, T. Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 Å resolution. *Nature* **391**, 251–258 (1998).
18. Artsimovitch, I. *et al.* Structural basis for transcription regulation by alarmone ppGpp. *Cell* **117**, 299–310 (2004).
19. Dieci, G. *et al.* A universally conserved region of the largest subunit participates in the active site of RNA polymerase III. *EMBO J.* **14**, 3766–3776 (1995).
20. Sosunov, V. *et al.* The involvement of the aspartate triad of the active center in all catalytic activities of multisubunit RNA polymerase. *Nucleic Acids Res.* **33**, 4202–4211 (2005).
21. Zaychikov, E. *et al.* Mapping of catalytic residues in the RNA polymerase active center. *Science* **273**, 107–109 (1996).
22. Langelier, M. F. *et al.* The highly conserved glutamic acid 791 of Rpb2 is involved in the binding of NTP and Mg(B) in the active center of human RNA polymerase II. *Nucleic Acids Res.* **33**, 2629–2639 (2005).
23. Lee, J., Kashlev, M., Borukhov, S. & Goldfarb, A. A β subunit mutation disrupting the catalytic function of *Escherichia coli* RNA polymerase. *Proc. Natl Acad. Sci. USA* **88**, 6018–6022 (1991).
24. Werner, F. & Weinzierl, R. O. A recombinant RNA polymerase II-like enzyme capable of promoter-specific transcription. *Mol. Cell* **10**, 635–646 (2002).
25. Holmes, S. F., Santangelo, T. J., Cunningham, C. K., Roberts, J. W. & Erie, D. A. Kinetic investigation of *Escherichia coli* RNA polymerase mutants that influence nucleotide discrimination and transcription fidelity. *J. Biol. Chem.* **281**, 18677–18683 (2006).
26. Temiakov, D. *et al.* Structural basis of transcription inhibition by antibiotic streptolydigin. *Mol. Cell* **19**, 655–666 (2005).
27. Tuske, S. *et al.* Inhibition of bacterial RNA polymerase by streptolydigin: stabilization of a straight-bridge-helix active-center conformation. *Cell* **122**, 541–552 (2005).
28. Artsimovitch, I., Svetlov, V., Murakami, K. S. & Landick, R. Co-overexpression of *Escherichia coli* RNA polymerase subunits allows isolation and analysis of mutant enzymes lacking lineage-specific sequence insertions. *J. Biol. Chem.* **278**, 12344–12355 (2003).
29. Vassilyev, D. G. *et al.* Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 Å resolution. *Nature* **417**, 712–719 (2002).
30. Abbondanzieri, E. A., Greenleaf, W. J., Shaevitz, J. W., Landick, R. & Block, S. M. Direct observation of base-pair stepping by RNA polymerase. *Nature* **438**, 460–465 (2005).
31. Huang, J., Briebe, L. G. & Sousa, R. Misincorporation by wild-type and mutant T7 RNA polymerases: identification of interactions that reduce misincorporation rates by stabilizing the catalytically incompetent open conformation. *Biochemistry* **39**, 11571–11580 (2000).
32. Laptchenko, O., Lee, J., Lomakin, I. & Borukhov, S. Transcript cleavage factors GreA and GreB act as transient catalytic components of RNA polymerase. *EMBO J.* **22**, 6322–6334 (2003).
33. Sosunova, E. *et al.* Donation of catalytic residues to RNA polymerase active center by transcription factor Gre. *Proc. Natl Acad. Sci. USA* **100**, 15469–15474 (2003).
34. Touloukhonov, I., Zhang, J., Palangat, M. & Landick, R. A central role of the RNA polymerase trigger loop in active-site rearrangement during transcriptional pausing. *Mol. Cell* (in the press).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank V. Svetlov for help with sequence alignments and critical reading of the manuscript. Use of the Advanced Photon Source was supported by the US Department of Energy, Office of Energy Research. This work was supported by grants from the NIH to D.G.V., I.A. and R.L.

Author Contributions D.G.V. determined and analysed the structure and supervised the project. M.N.V. performed crystallization and data collection. J.Z. and M.P. performed biochemical analysis of the trigger loop mutants under the guidance of R.L. I.A. performed biochemical analysis of the *T. thermophilus* TECs. D.G.V., I.A. and R.L. jointly wrote the manuscript.

Author Information The atomic coordinates and structure factors of the tTEC/AMPcPP and tTEC/AMPcPP/Stl complexes are deposited in the Protein Data Bank under accession numbers 2O5J and 2PPB, respectively. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to D.G.V. (dmitry@uab.edu).

METHODS

Crystallization and data collection. To obtain the crystals of the tTEC/AMPcPP and tTEC/AMPcPP/Stl complexes, the tTEC prepared in the same way as described previously¹⁰ was co-crystallized with AMPcPP (3 mM) or with AMPcPP (3 mM) and Stl (3 mM), respectively, under essentially the same crystallization conditions (the well solution contained 30% 2-methyl-2,4-pentandiol, 0.1 M sodium cacodylate pH 6.5, 0.2 M magnesium acetate) as the ligand-free tTEC³⁵. The crystallization trials yielded the crystals of the same size and shape as for the ligand-free tTEC that belong to the same $P4_12_12$ space group. Whereas the crystals of the tTEC/AMPcPP/Stl complex had nearly the same unit cell dimensions ($a = b = 155.38 \text{ \AA}$, $c = 496.99 \text{ \AA}$) as those of the ligand-free tTEC, the crystals of the tTEC/AMPcPP complex were characterized by a large change in the unit cell parameters ($a = b = 152.38 \text{ \AA}$, $c = 524.57 \text{ \AA}$) (Supplementary Table S1). The crystals of both complexes were improved and selected for data collection with the same strategy as for the ligand-free complex¹⁰. This allowed us finally to select a few well-diffracting crystals, from which the high-resolution data were collected at the SERCAT beam line (APS, Argonne, USA; Supplementary Table S1). The structure determination and refinement were performed with the same approaches and procedures as for the ligand-free tTEC¹⁰ (Supplementary Table S1).

Determination of the structures. The tTEC/AMPcPP/Stl crystals were characterized by essentially the same unit cell parameters as those of the ligand-free tTEC. In contrast, the tTEC/AMPcPP crystals consistently demonstrated a marked increase ($\sim 25 \text{ \AA}$) in the dimensions of the c axis (Supplementary Table S1). Given that all the crystals were obtained under the same conditions, this result suggests significant alterations of the enzyme structure induced by substrate binding, probably blocked in the presence of Stl, and incompatible with the previous crystal form. Indeed, the application of zonal scaling to the tTEC/AMPcPP/Stl data in the $P4_12_12$ space group yielded an initial R -factor/ R_{free} ratio of $\sim 40\%$ that decreased to 38.5% at 3.0 \AA resolution on domain rigid-body refinement, a value very similar to that of the tTEC at the same stage of refinement¹⁰. In contrast, the rigid-body refinement did not converge for the EC/AMPcPP structure, whereas molecular replacement with the program AMORE³⁶ revealed a $\sim 7.5^\circ$ rotation of the entire molecule, giving an initial R -factor/ R_{free} of $\sim 42\%$. The domain rigid-body refinement³⁶ further improved the R -factor by $\sim 4\%$ (38.2%), indicating more extensive domain rearrangement in the AMPcPP-bound EC than in the tTEC/AMPcPP/Stl structure. The twinning option of the CNS program³⁷ allowed us to improve the ED and the R -factor values further (Supplementary Table S1).

The focal point in both complexes is the active site area, in which local but functionally significant alterations may occur on substrate binding. To detect these putative alterations and to minimize the model bias, the first base pair of the RNA/DNA hybrid (position -1) and the acceptor template nucleotide (nt, register $i + 1$) were omitted from the model at the initial stages of refinement. The first base pair of the dwDNA (register $i + 2$) was also omitted: if the 'multi-substrate' hypothesis⁹ were correct, the substrate would disrupt the $(i + 2)$ base pair because the scaffold contained Ts in the template strand at both $(i + 1)$ and $(i + 2)$. The resulting electron density (ED) map showed clear ED and allowed us to model³⁸ all the missing portions of the structure, including the $(i + 2)$ base pair¹⁰ arguing against the 'multisubstrate' model. The refinement converged to the final R -factor/ R_{free} values of $0.225/0.257$ and $0.234/0.266$ for the tTEC/AMPcPP and tTEC/AMPcPP/Stl structures, respectively (Fig. 1 and Supplementary Table S1). All structural figures were prepared with the programs Molscript, Bobscript and Raster3D^{39–41}.

Isolation and assay of mutant *E. coli* RNAPs. Core wild-type and Δ SI3 RNAPs were purified after overexpression from plasmids as described previously²⁸. PP and AA mutant RNAPs were obtained by oligonucleotide mutagenesis of plasmid pRL662, transfer of a *BsmI*–*XhoI* fragment to pRL4455 (ref. 34), and purification after overexpression by polyethyleneimine precipitation, Ni^{2+} -nitrilotriacetate chromatography and heparin-agarose chromatography. TECs were reconstituted as described⁴², using for A14 TEC (nucleotide addition assay) oligonucleotides 5'-CTCTGAATCTCTCCAGCACACATTAGGACGTACTG-ACC-3' (template DNA), 5'-biotin-GGTCAGTACGTCCATTTCGATCTCCC-GAAGAGATTCAGAG-3' (non-template DNA) and 5'-AGGCGAUGUGUG-3' (RNA). The reconstituted TECs were 3'-labelled by incubation with $2 \mu\text{M}$ [α -³²P]ATP (A14) or UTPs (each at $10 \mu\text{M}$) and [α -³²P]CTP. For elongation assays, A14 TECs were diluted in buffer A (25 mM HEPES-KOH pH 8.0, 130 mM KCl, 5 mM MgCl_2 , 1 mM dithiothreitol, 0.15 mM EDTA, $25 \mu\text{g ml}^{-1}$ acetylated BSA) to 25–50 nM TEC, and assayed at various concentrations of CTP at 37°C with a KinTek rapid-quench flow apparatus. The concentration of MgCl_2 was adjusted to maintain Mg^{2+} in 5 mM excess of the CTP concentration. Samples were quenched by addition of 1 vol. of 2 M HCl, immediately neutralized to pH 8.0 by addition of the same volume of 3 M unbuffered Tris, extracted with phenol, precipitated with ethanol, and suspended in formamide loading dye. For pyrophosphorolysis assays, U14 TECs were immobilized on streptavidin-agarose

beads (Sigma), washed extensively with buffer B (10 mM Tris-HCl pH 7.9, 20 mM NaCl, 10 mM MgCl_2 , 14 mM 2-mercaptoethanol, 0.1 mM EDTA), incubated with PP_i (1 mM) in buffer B for 10 min at 37°C , and dissociated by heating at 94°C for 3 min after the addition of an equal volume of saturated urea in 90 mM Tris-borate pH 8.3, 50 mM EDTA to halt the reactions. Reaction products were separated electrophoretically on denaturing polyacrylamide gels (7 M urea, 17.5% acrylamide:bisacrylamide, 19:1), and RNA quantities were determined from Phosphorimager scans of the gels.

Reaction rates (addition of CMP to A14 ECs and pyrophosphorolysis of U14 ECs; Fig. 4d and Supplementary Fig. 10) were determined as described⁸ by non-linear regression of RNA concentrations against time with errors calculated from the standard error of regression and Student's t distribution at the 95% confidence level. Apparent V_{max} and $\text{CTP}_{0.5}$ values (Fig. 4d) were calculated by non-linear regression of the average rate weighted by their respective error estimates against CTP concentration assuming a hyperbolic dependence, with error calculated as described for the rate measurements.

35. Kashkina, E. *et al.* Elongation complexes of *Thermus thermophilus* RNA polymerase that possess distinct translocation conformations. *Nucleic Acids Res.* **34**, 4036–4045 (2006).
36. Navaza, J. Implementation of molecular replacement in AMoRe. *Acta Crystallogr. D Biol. Crystallogr.* **57**, 1367–1372 (2001).
37. Brunger, A. T. *et al.* Crystallography and NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.* **54**, 905–921 (1998).
38. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
39. Kraulis, P. J. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Cryst.* **24**, 946–950 (1991).
40. Esnouf, R. M. Further additions to MolScript version 1.4, including reading and contouring of electron-density maps. *Acta Crystallogr. D Biol. Crystallogr.* **55**, 938–940 (1999).
41. Merrit, E. A. & Bacon, D. J. Raster3D: photorealistic molecular graphics. *Methods Enzymol.* **277**, 505–524 (1997).
42. Sidorenkov, I., Komissarova, N. & Kashlev, M. Crucial role of the RNA:DNA hybrid in the processivity of transcription. *Mol. Cell* **2**, 55–64 (1998).

Water vapour in the atmosphere of a transiting extrasolar planet

Giovanna Tinetti^{1,2,3}, Alfred Vidal-Madjar³, Mao-Chang Liang^{4,5}, Jean-Philippe Beaulieu³, Yuk Yung⁵, Sean Carey⁶, Robert J. Barber², Jonathan Tennyson², Ignasi Ribas⁷, Nicole Allard³, Gilda E. Ballester⁸, David K. Sing^{3,9} & Franck Selsis¹⁰

Water is predicted to be among the most abundant (if not the most abundant) molecular species after hydrogen in the atmospheres of close-in extrasolar giant planets ('hot Jupiters')^{1,2}. Several attempts have been made to detect water on such planets, but have either failed to find compelling evidence for it^{3,4} or led to claims that should be taken with caution⁵. Here we report an analysis of recent observations of the hot Jupiter HD 189733b (ref. 6) taken during the transit, when the planet passed in front of its parent star. We find that absorption by water vapour is the most likely cause of the wavelength-dependent variations in the effective radius of the planet at the infrared wavelengths 3.6 μm , 5.8 μm (both ref. 7) and 8 μm (ref. 8). The larger effective radius observed at visible wavelengths⁹ may arise from either stellar variability or the presence of clouds/hazes. We explain the report of a non-detection of water on HD 189733b (ref. 4) as being a consequence of the nearly isothermal vertical profile of the planet's atmosphere.

Water absorbs over a broad wavelength range, covering most of the infrared region and part of the visible region, and has a very distinctive spectral signature. Theoretical models anticipated that it could be detected on hot Jupiters by observing these planets during their primary transit (when they pass in front of their parent star) and estimating the effective radius of the planet at multiple wavelengths. Variations in this effective radius can be used to characterize the transmission spectrum of the planet's atmosphere^{1,10,11}. This technique had been successfully used to probe other atmospheric constituents on hot Jupiters^{12–15}. Recently water detection was claimed⁵ in the atmosphere of the hot Jupiter HD 209458b. Although it is possible that the observed variations in the effective radius of the planet in the visible/near-infrared ($\sim 0.9\text{--}1\ \mu\text{m}$)¹⁶ are due to the absorption of water¹⁰, the analysis⁵ was based on the very lowest-flux part of the spectrum, at the edge of the detector array where the largest systematic effects occur (figure 3 in ref. 16) and the overall noise is much larger. The quoted errors in the effective radius are based on photon noise alone, and do not include these systematic uncertainties¹⁶. These considerations alone are a good reason for caution.

Very recently, the planet-to-star radius ratios of another hot Jupiter, HD 189733b (ref. 6), were measured in three of the Spitzer Space Telescope Infrared Array Camera (IRAC) bands centred at 3.6 μm (ref. 7), 5.8 μm (ref. 7) and 8 μm (ref. 8). The corresponding transit depths, that is, the ratio of the projected area of the planet to that of the star, were estimated to be $2.356 \pm 0.02\%$ (ref. 7), $2.436 \pm 0.02\%$ (ref. 7) and $2.39 \pm 0.02\%$ (ref. 8). To interpret these data, we simulated transmission spectra, which determine the effective radius of the planet at different wavelengths, by improving

upon a previous spectral/planetary model¹¹. The radiative transfer and geometry are unchanged from that model: we use 45 atmospheric layers to describe the variation with altitude of the temperature, pressure (from 10 bar to $\sim 10^{-10}$ bar, corresponding to $\sim (4\text{--}6) \times 10^3$ km altitude, depending on the thermal profile), density and mixing ratios. The water volume mixing ratio is about 5×10^{-4} , assuming an elementary C/O ratio equal to the solar value². Studies of the sensitivity of these results to changes of water abundances are presented. The crucial difference from the previous model is the use of a new high-accuracy computed list of water lines, BT2 (ref. 17). Whereas only about 80,000 water line strengths are known experimentally, the BT2 line list contains more than half a billion transitions; these extra transitions become increasingly important at higher temperatures. As a result, the BT2 opacity at a particular wavelength is far more temperature-dependent than might be expected from computations using only experimental data. The opacities were calculated for the selected spectral band at different temperatures from 500 K to 2,000 K, and interpolated for intermediate values of the temperature at each atmospheric layer.

We repeated our simulations with different thermal profiles, compatible with three-dimensional climate models of hot Jupiters for the day/night sides and morning/evening terminators¹⁸. These simulations predict adiabatic profiles colder at the μbar level ($\sim 500\text{--}700$ K) and warmer at 1–10 bar level ($\sim 1,500$ K) for the nightside and the morning/evening terminators; and they predict nearly isothermal profiles in the upper ($\sim 800\text{--}1,000$ K) and lower part of the atmosphere ($\sim 2,000$ K) for the dayside. The temperature might increase again ($\sim 2,000$ K) at pressures $\sim 10^{-6}\text{--}10^{-10}$ bar. This increase is consistent with both models and observations of the planets in our Solar System and beyond¹⁵. We used photochemical models^{2,19} with updated nitrogen chemistry from the NIST database (<http://www.nist.gov/srd/chemkin.htm>) to estimate the ammonia abundance. In general, the predicted ammonia mixing ratios are $\ll 10^{-7}$. The ammonia absorption coefficients were estimated using HITRAN²⁰ data corrected for the higher temperatures. This additional absorption by ammonia makes negligible contribution to the observed infrared absorptions. CO and CH₄ also absorb in the spectral range considered. CO, if present in sufficient abundance to be detected², should show its signature in the 4.5 μm IRAC band¹¹, not yet used for observations of primary transit. CH₄ is ruled out both by the relative contribution of the observed IRAC bands and by the predictions of the photochemical models².

In our simulations, sodium and potassium were included with solar abundances. The line shapes of these alkali metals were calculated at

¹European Space Agency, Frascati 00044, Italy. ²Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, UK. ³Institut d'Astrophysique de Paris, CNRS, Université Pierre et Marie Curie, 75014 Paris, France. ⁴Research Center for Environmental Changes, Academia Sinica, Taipei 115, Taiwan. ⁵Division of Geological and Planetary Sciences, ⁶IPAC-Spitzer Science Center, California Institute of Technology, Pasadena, California 91125, USA. ⁷Institut de Ciències de l'Espai (CSIC-IEEC), Campus UAB, 08193 Bellaterra, Spain. ⁸Department of Planetary Sciences, Lunar and Planetary Laboratory, University of Arizona, Tucson, 85721-0063, Arizona, USA. ⁹Centre National d'Etudes Spatiales, Paris 75001, France. ¹⁰Ecole Normale Supérieure, CNRS, 69364 Lyon, France.

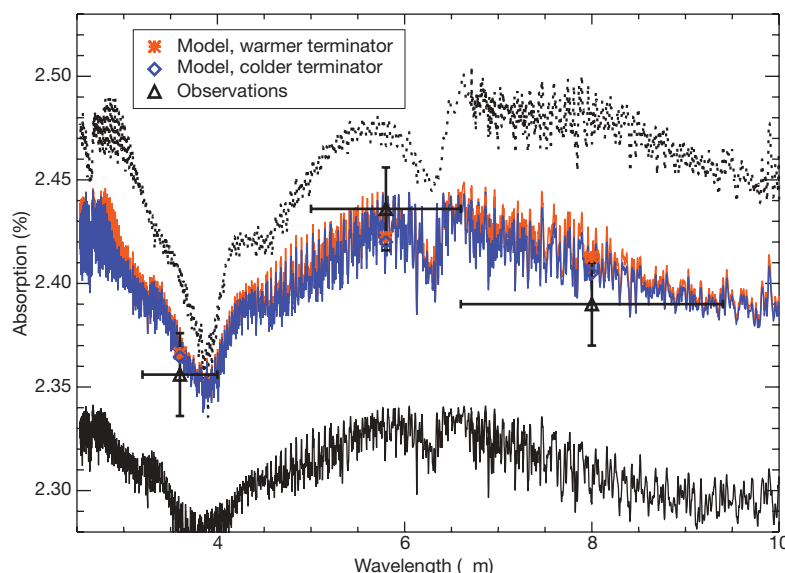


Figure 1 | A comparison of the observations with simulated water absorption. In these simulated transmission spectra, the water mixing ratio profile is assumed to be constant² and equal to $\sim 5 \times 10^{-4}$. The observations are indicated with black triangles and error bars at 1σ , and the coloured rhombi and stars indicate the different models integrated over the IRAC bands. To match the observations, the planetary radius at 10 bar corresponds to a transit depth of 2.28%. Blue trace, colder terminator¹⁸ temperature–pressure profile; orange trace, warmer terminator¹⁸; black traces, constant temperature at 500 K (solid line) and 2,000 K (dotted line). Warmer temperatures increase the atmospheric scale-height (that is, the vertical distance over which the pressure decreases by a factor of e), hence the atmosphere is optically thick at higher altitudes. This explains the differences among the three classes of spectra at wavelengths shorter than

different temperatures and interpolated for intermediate values^{21,22}. Their spectral contribution becomes important in the visible wavelength range (see below). The H_2 – H_2 opacity²³ was interpolated to the temperature of each atmospheric layer. As collision-induced absorption scales with the square of the pressure, the H_2 – H_2 contribution becomes important for pressures higher than ~ 1 bar.

The difference in absorption depths found in the IRAC bands centred at 3.6, 5.8 and 8 μm can be explained by the presence of water vapour in the atmosphere of HD 189733b. Figure 1 compares our calculated water absorption with the observations. The sensitivity of our simulations to temperature and water abundance is also considered in the figure. The scenarios that match the observations best are terminator profiles¹⁸. These profiles (see bold curves in figure 4 in ref. 18) are in agreement with recent estimates of the hemisphere-averaged brightness temperature⁸.

We extended our simulations of the transmission spectra from the infrared to the visible to check the consistency of the current interpretation with the observations at shorter wavelengths⁹. The measure of the transit depth in the optical was found to be $2.48 \pm 0.05\%$ (ref. 9), higher than the infrared values. The extra absorption in the visible is probably due to the effect of star spots⁹ and/or the presence of optically thick clouds/hazes in the visible wavelength range (Fig. 2). We note that large star spots or a condensate/haze where the particle size is less than 1 μm may not affect the infrared transmission spectrum^{10,24}. This scenario is in agreement with models and observations of the planet HD 209458b, which has characteristics similar to the planet HD 189733b considered here¹⁰. For HD 209458b, the transit depth at 24 μm (ref. 25) does not differ significantly from the visible results. This observational constraint is consistent with our simulations for HD 189733b (Fig. 2). Our explanation differs from a recent paper that proposes a cloud-free atmosphere with rainout and photo-ionization to fit multiband photometry measurements of HD 209458b in the optical¹⁵.

$\sim 3.5 \mu m$ and longer than $\sim 4.5 \mu m$, where the water opacities are far less temperature-dependent. The opposite is true for the water opacities in the 3.5–4.5 μm wavelength range, which might be orders of magnitude smaller at 500 K rather than at 2,000 K, so for colder temperature profiles the weaker water lines are optically thick at ~ 10 bar or deeper. An increase/decrease of the mixing ratio by a factor of 10 with respect to the standard case considered will cause, as a main effect, an increase/decrease of ~ 0.03 – 0.04% in the total absorption due to water. As a secondary effect, the absorption gradient between 3.6 and 5.8 μm gets steeper for lower water mixing ratios, but this trend is marginal compared to the role played by temperature. CO, if present in sufficient abundance to be detected², would show its distinctive signature in the 4.5–4.9 μm spectral range (see ref. 11 for details). This is a spectral region than can be observed with IRAC (channel centred at 4.5 μm).

Finally, we simulated the emission spectra of HD 189733b with the same atmospheric constituents but different thermal profiles. These synthetic spectra are useful for comparing our analysis of the primary transit observations—which indicate the presence of water vapour in the atmosphere of HD 189733b—with the most recent secondary transit observations of HD 189733b and HD 209458b made with

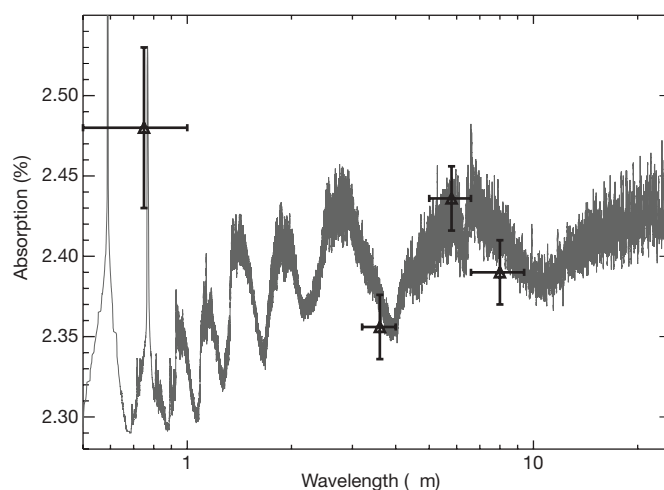


Figure 2 | A comparison of the observations of primary transit with a simulated infrared and optical transmission spectrum. The transmission spectrum is here modelled from 0.5 to 25 μm (grey trace). We assumed a water mixing ratio of $\sim 5 \times 10^{-4}$ and a cold terminator temperature–pressure profile. The observations are indicated with black triangles and error bars at 1σ . Absorptions due to alkali metals and water are unlikely to be the cause of the extra absorption observed at visible wavelengths: the mean absorption over the 0.5–1 μm band is significantly less than the measured value. Star spots or optically thick condensates/hazes would be a good explanation.

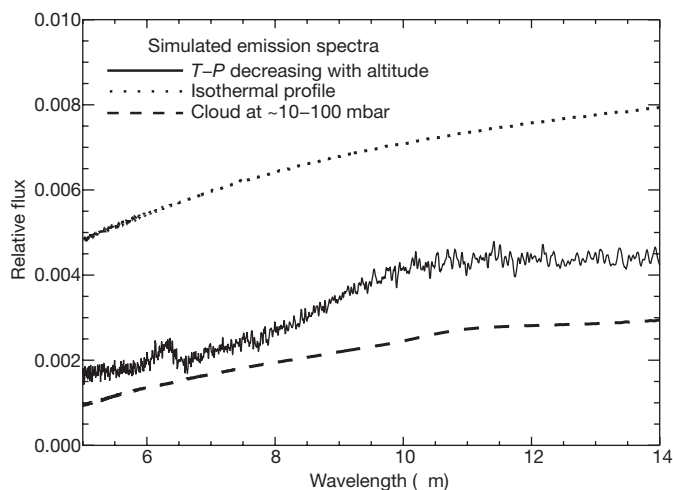


Figure 3 | Simulated emission spectra of HD 189733b in the infrared. We show here different simulated emission spectra, all generated at the same spectral resolution and corresponding to the same atmospheric composition but different temperature–pressure (T – P) profiles. Accordingly, we obtain very different spectral responses: hot isothermal profile (top, dotted line), isothermal profile and a cloud at ~ 10 – 100 mbar altitude (bottom, dashed), and temperature profile decreasing with altitude (middle, solid). Although present and with the same abundances in all the three scenarios, water is detectable through emission spectra only in the last case, showing up as a steep gradient between ~ 8 and $10\ \mu\text{m}$ when compared to the 10 – $14\ \mu\text{m}$ region. An isothermal profile in the upper part^{18,28} and a cloud in the lower part of the atmosphere is sufficient to explain the non-detection of water⁴ even if water were present in high abundance. HD 189733b is presumably tidally locked, therefore the thermal profiles and the condensate dynamics might be very different on the two sides of the planet. The same reasoning applies to HD 209458b and the most recent observations in secondary transit of that planet³.

the Spitzer Infrared Spectrograph (IRS)^{3,4}, which failed to detect water in the infrared emission spectra in the 7.5 – $14\ \mu\text{m}$ region. Secondary transit is another technique to study the characteristics of hot Jupiters' atmospheres. This method involves collecting the photons emitted directly by the planet and measuring the brightness of the star plus planet system and its progressive dimming when the planet is occulted by the parent star^{26,27}. The previous non-detection of water on HD 189733b (ref. 4) does not contradict the interpretation we present here. Secondary transit observations require a significant temperature gradient to be sensitive to atmospheric molecules absorbing in the infrared, but strong circulation on hot Jupiters can flatten the dayside temperature gradient²⁸. This phenomenon is illustrated in a classic paper²⁹ that gives the first portrait of the Earth in the thermal infrared from space: despite the fact that CO_2 is nearly uniform in the atmosphere, the $667\ \text{cm}^{-1}$ CO_2 feature is prominent in the tropics, where the thermal gradient between the surface and the upper atmosphere is large ($\sim 100\ \text{K}$), but it is hardly detectable in polar observations, where stratospheric and surface temperatures are nearly the same²⁴. A simple model of HD 189733b infrared emission spectra illustrating this effect is given in Fig. 3. We note a recent different explanation³⁰: the inconsistency of the spectral data for HD 189733b obtained with IRS⁴ with the $8\ \mu\text{m}$ photometry obtained with IRAC⁸ may indicate that the IRS spectra from 7.5 to $10\ \mu\text{m}$ are not completely reliable.

Our result provides the first detection of a molecular species in an exoplanet atmosphere in the infrared. Although the detection of water is secure, more observations at multiple wavelengths are needed to constrain its abundance. Moreover, the IRAC band centred at $4.5\ \mu\text{m}$ could constrain the presence of CO on HD 189733b, and indirectly the C/O ratio¹¹. Additional observations in secondary transit in the infrared will help to refine the thermal profiles, and give more information on the presence of condensates/hazes and their variability with time on the dayside. Analogous observations are desirable for other hot Jupiters to start comparative planetology for this class of objects.

Received 8 April; accepted 1 June 2007.

- Seager, S. & Sasselov, D. D. Theoretical transmission spectra during extrasolar giant planet transits. *Astrophys. J.* **537**, L916–L921 (2000).
- Liang, M. C., Parkinson, C. D., Lee, A. Y.-T., Yung, Y. L. & Seager, S. Source of atomic hydrogen in the atmosphere of HD209458b. *Astrophys. J.* **596**, L247–L250 (2003).
- Richardson, L. J., Deming, D., Horning, K., Seager, S. & Harrington, J. A spectrum of an extrasolar planet. *Nature* **445**, 892–895 (2007).
- Grillmair, C. J. *et al.* A Spitzer spectrum of the exoplanet HD189733b. *Astrophys. J.* **658**, L115–L118 (2007).
- Barman, T. Identification of absorption features in an extrasolar planet atmosphere. *Astrophys. J.* **661**, L191–L194 (2007).
- Bouchy, F. *et al.* ELODIE metallicity-biased search for transiting hot Jupiters II. A very hot Jupiter transiting the bright K star HD189733. *Astron. Astrophys.* **444**, L15–L19 (2005).
- Beaulieu, J. P. *et al.* Spitzer observations of the primary transit of the planet HD189733b at 3.6 and $5.8\ \mu\text{m}$. *Astrophys. J.* (submitted).
- Knutson, H. A. *et al.* A map of the day–night contrast of the extrasolar planet HD 189733b. *Nature* **447**, 183–186 (2007).
- Winn, J. N. *et al.* The transit light curve project. V. System parameters and stellar rotation period of HD 189733. *Astron. J.* **133**, 1828–1835 (2007).
- Brown, T. M. Transmission spectra as diagnostics of extrasolar giant planet atmospheres. *Astrophys. J.* **553**, 1006–1026 (2001).
- Tinetti, G. *et al.* Infrared transmission spectra for extrasolar giant planets. *Astrophys. J.* **654**, L99–L102 (2007).
- Charbonneau, D., Brown, T. M., Noyes, R. W. & Gilliland, R. L. Detection of an extrasolar planet atmosphere. *Astrophys. J.* **568**, 377–384 (2002).
- Vidal-Madjar, A. *et al.* An extended upper atmosphere around the extrasolar planet HD209458b. *Nature* **422**, 143–146 (2003).
- Vidal-Madjar, A. *et al.* Detection of oxygen and carbon in the upper atmosphere of the extrasolar planet HD209458b. *Astrophys. J.* **604**, L69–L72 (2004).
- Ballester, G. E., Sing, D. K. & Herbert, F. The signature of hot hydrogen in the atmosphere of the extrasolar planet HD 209458b. *Nature* **445**, 511–514 (2007).
- Knutson, H. A., Charbonneau, D., Noyes, R. W., Brown, T. M. & Gilliland, R. L. Using stellar limb-darkening to refine the properties of HD 209458b. *Astrophys. J.* **655**, 564–575 (2007).
- Barber, R. J., Tennyson, J., Harris, G. J. & Tolchenov, R. A high accuracy computed water line list. *Mon. Not. R. Astron. Soc.* **368**, 1087–1094 (2006).
- Burrows, A., Sudarsky, D. & Hubeny, I. Theory for the secondary eclipse fluxes, spectra, atmospheres and light curves of transiting extrasolar giant planets. *Astrophys. J.* **650**, 1140–1149 (2006).
- Liang, M. C., Yung, Y. L. & Shemansky, D. E. Photolytically generated aerosols in the mesosphere and thermosphere of Titan. *Astrophys. J.* **661**, L199–L201 (2007).
- Rothman, L. S. *et al.* The HITRAN 2004 molecular spectroscopic database. *J. Quant. Spectrosc. Radiat. Transf.* **96**, 139–204 (2005).
- Allard, N. F., Allard, F., Hauschildt, P. H., Kielkopf, J. F. & Machin, L. A new model for brown dwarf spectra including accurate unified line shape theory for the Na I and K I resonance line profiles. *Astron. Astrophys.* **411**, L473–L476 (2003).
- Allard, N. F., Spiegelman, F. & Kielkopf, J. F. Study of the K–H₂ quasi molecular line satellite in the potassium resonance line. *Astron. Astrophys.* **465**, 1085–1091 (2007).
- Borysow, A., Jorgensen, U. G. & Fu, Y. High temperature (1000–7000K) collision induced absorption of H₂ pairs computed from the first principles, with application to cool and dense stellar atmospheres. *J. Quant. Spectrosc. Radiat. Transf.* **68**, 235–255 (2001).
- Goody, R. M. & Yung, Y. L. *Atmospheric Radiation* (Oxford Univ. Press, New York, 1989).
- Richardson, L. J., Harrington, J., Seager, S. & Deming, D. A Spitzer infrared radius for the transiting extrasolar planet HD209458b. *Astrophys. J.* **649**, 1043–1047 (2006).
- Charbonneau, D. *et al.* Detection of thermal emission from an extrasolar planet. *Astrophys. J.* **626**, 523–529 (2005).
- Deming, D., Seager, S., Richardson, L. J. & Harrington, J. Infrared radiation from an extrasolar planet. *Nature* **434**, 740–743 (2005).
- Fortney, J. J., Cooper, C. S., Showman, A. P., Marley, M. S. & Freedman, R. S. The influence of atmospheric dynamics on the infrared spectra and light curves of hot Jupiters. *Astrophys. J.* **652**, 746–757 (2006).
- Hanel, R. A. *et al.* The Nimbus 4 infrared spectroscopy experiment 1. Calibrated thermal emission spectra. *J. Geophys. Res.* **77**, 2629–2639 (1972).
- Fortney, J. J. & Marley, M. S. Analysis of Spitzer mid infrared spectra of irradiated planets: Evidence for water vapor? *Astrophys. J.* (submitted).

Acknowledgements We thank A. Lecavelier, D. Ehrenreich, J.-M. Désert, Roger Ferlet and G. Hebrard for their work on the IRAC observations. We thank A. Noriega-Crespo and the Spitzer Staff for helping to schedule the observations with IRAC, and E. Lellouch, A. Morbidelli, B. Schultz, F. Bouchy and J. B. Marquette for useful input to the paper. M.-C.L. and Y.Y. were supported by NASA and the Virtual Planetary Laboratory at the California Institute of Technology.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to G.T. (gio@gps.caltech.edu).

LETTERS

Compositional homogeneity in the fragmented comet 73P/Schwassmann–Wachmann 3

N. Dello Russo¹, R. J. Vervack Jr¹, H. A. Weaver¹, N. Biver², D. Bockelée-Morvan², J. Crovisier² & C. M. Lisse¹

The remarkable compositional diversity of volatile ices within comets^{1–3} can plausibly be attributed to several factors, including differences in the chemical, thermal and radiation environments in comet-forming regions, chemical evolution during their long storage in reservoirs far from the Sun⁴, and thermal processing by the Sun after removal from these reservoirs. To determine the relevance of these factors, measurements of the chemistry as a function of depth in cometary nuclei are critical. Fragmenting comets expose formerly buried material, but observational constraints have in the past limited the ability to assess the importance of formative conditions and the effects of evolutionary processes on measured composition^{5–8}. Here we report the chemical composition of two distinct fragments of 73P/Schwassmann–Wachmann 3. The fragments are remarkably similar in composition, in marked contrast to the chemical diversity within the overall comet population and contrary to the expectation that short-period comets should show strong compositional variation with depth in the nucleus owing to evolutionary processing from numerous close passages to the Sun. Comet 73P/Schwassmann–Wachmann 3 is also depleted in the most volatile ices compared to other comets, suggesting that the depleted carbon-chain chemistry seen in some comets from the Kuiper belt reservoir is primordial and not evolutionary¹.

Comet 73P/Schwassmann–Wachmann 3 (hereafter 73P) is a Jupiter-family comet that probably originated in the Kuiper belt, on the basis of the short period (5.34 yr) and low inclination (11.4°) of its orbit. Narrowband photometric observations during the 1990 and 2006 apparitions demonstrated that 73P is depleted in C₂ relative to CN and OH (refs 9, 10), placing it in the ‘carbon-chain depleted’ class of comets¹. During its 1995 apparition, the comet split into at least five fragments¹¹, increasing the overall gas production rate by more than an order of magnitude from the previous apparition¹². Observations obtained during the 2006 apparition confirm the disintegration of 73P: 68 named fragments were identified, and at least two of the larger fragments were shedding boulder-sized pieces¹³. Radar observations confirm that the largest fragments (C and B) are of significant size (hundreds of metres in diameter)¹⁴, strongly suggesting that material from different depths (including the deep interior of 73P) is exposed on the surfaces of these bodies.

Jupiter-family comets, although essential for creating a complete chemical taxonomy of comets, are generally too faint for detailed spectroscopic analysis. Thus, they are underrepresented in current surveys, both in number of comets and in detected parent volatiles (species released directly from ices in the nucleus and not products of chemistry in the coma), relative to long-period comets. 73P made a very close approach to Earth in 2006 (Supplementary Table S1), allowing the parent volatile inventory of individual pieces of a fragmenting comet to be determined with unprecedented detail, and

providing a sensitive test for chemical heterogeneity within a comet nucleus.

We measured the volatile composition of the B and C fragments of comet 73P using the Cryogenic Echelle Spectrometer (CSHELL) at the NASA Infrared Telescope Facility¹⁵ (IRTF) and the Near-Infrared Spectrometer (NIRSPEC) at the Keck II telescope¹⁶ on Mauna Kea, Hawaii (Supplementary Table S1). Spectra were acquired between wavelengths (λ) of 2.8 and 4.7 μm at a spectral resolving power sufficient to detect the diagnostic rotational-vibrational lines of parent volatiles in the coma ($\lambda/\Delta\lambda \approx 25,000$). Details of our observing strategy and data processing for both CSHELL and NIRSPEC observations have been given elsewhere^{17,18}. Sample spectra, with the best-fit atmospheric models superimposed, are shown in Fig. 1 (see also Supplementary Figs S1 and S2).

In order to obtain quantitative information from flux-calibrated spectra, the temperature-dependent fluorescence efficiencies (g -factors) for individual rotational-vibrational lines are needed. Fluorescence models have been developed for all sampled parent volatiles^{19–21} (see also Supplementary Information). The measurement of multiple lines with a wide range of ground-state rotational energies enabled the use of these models to determine rotational temperatures for some species (Table 1, Supplementary Fig. S3) using established methodologies¹⁹; otherwise, a reasonable rotational temperature was adopted (Table 1). A sufficient number of both ortho and para H₂O lines were detected to determine ortho-to-para ratios (OPRs) and associated nuclear spin temperatures for both fragments on UT 2006 May 14 and 15. The significance of OPRs is still uncertain, but they may preserve information about conditions in comet-forming regions in the early solar nebula (Supplementary Information). Measurements on both fragments are consistent with the high-temperature statistical equilibrium value of OPR = 3 and spin temperatures greater than 42 and 37 K for 73P-B and 73P-C, respectively (Table 1; see also Supplementary Information).

Integrated column densities for individual lines were extracted using g -factors determined at the measured or adopted rotational temperatures and OPRs, allowing the determination of molecular production rates and relative abundances in 73P-B and 73P-C (Table 1). As might be expected for a fragmenting comet, significant day-to-day variation in absolute production rates was observed, especially for 73P-B. Despite these daily fluctuations in absolute production rates, relative abundances within each fragment remained consistent (Table 1, Fig. 2a). Furthermore, the relative abundances between fragments B and C are remarkably similar to each other when compared to the diversity in chemistry within the overall comet population (Fig. 2b). Given that we are sampling the interior of a recently disrupted comet¹¹, in which layers of increasing depth within the original intact nucleus have been revealed, these results provide strong evidence that the nucleus of 73P is chemically homogeneous

¹Space Department, The Johns Hopkins University Applied Physics Laboratory, 11100 Johns Hopkins Road, Laurel, Maryland 20723-6099, USA. ²Observatoire de Meudon–Paris, LESIA, 5 Place Jules Janssen, Meudon 92190, France.

and that its composition primarily reflects formative conditions and not evolutionary processing.

Earlier infrared observations of 73P acquired in April 2006 reported detections of H₂O in the B and C fragments, and HCN, C₂H₆ and C₂H₂ in the C fragment⁸. Relative abundances in the C fragment from that work agree with those reported here except for C₂H₂. This species was claimed to be several times more abundant than measured upper limits from the present work, and in the normal range of abundances found in the overall comet population⁸—a surprising result, considering that multiple measurements have shown a consistent C₂ depletion in 73P (refs 9, 10). This earlier result might suggest chemical heterogeneity in the nucleus of 73P; however, we regard the detection as unconvincing owing to the relatively low quality of the data and the marginal nature of the reported C₂H₂ detection (Supplementary Information)⁸.

We note that the diversity of formative conditions and processing histories experienced by comets makes it unlikely that the

homogeneous and primitive composition of 73P is representative of all comets. Molecular abundances, presumably from the interior of comet nuclei, have been determined for several fragmenting comets (Supplementary Information). In the case of comet C/2001 A2 (LINEAR), evidence suggests chemical heterogeneity within the nucleus⁷, whereas no conclusive evidence for heterogeneity has been found in other fragmenting comets^{5,6}. Given that 73P and C/2001 A2 (LINEAR) are compositionally different⁷ and are members of

Table 1 | Volatile production rates, abundances, rotational temperatures and OPRs in 73P

Species	UT date (May 2006)	T_{rot}^* (K)	OPR*	T_{spin}^\dagger (K)	Production rate* (10^{25} s^{-1})	Abundance*, H ₂ O = 100
73P/Schwassmann–Wachmann 3-B						
H ₂ O	9.5	120	3.0	>50	3,483 ± 290	
H ₂ O	14.5–14.6	110 ⁺⁵ ₋₃	3.2 ± 0.3	>42	1,208 ± 88	
H ₂ O	15.6	98 ⁺⁴ ₋₃	3.2 ± 0.3	>42	502 ± 29	
CO	9.5	100			<67 (3σ)	<1.9 (3σ)
CH ₃ OH	9.5	100			7.17 ± 1.31	0.206 ± 0.041
CH ₃ OH	14.5	90			2.14 ± 0.29	0.177 ± 0.025
CH ₃ OH	15.6	80			1.70 ± 0.33	0.339 ± 0.066
HCN	9.5	100			8.78 ± 0.80	0.252 ± 0.031
HCN	14.6	87 ⁺⁵ ₋₄			3.37 ± 0.34	0.279 ± 0.019
HCN	15.6	73 ± 3			1.44 ± 0.14	0.287 ± 0.015
C ₂ H ₆	9.4	100			4.74 ± 0.54	0.136 ± 0.019
C ₂ H ₆	14.5	127 ⁺⁴⁰ ₋₃₀			1.88 ± 0.28	0.156 ± 0.024
C ₂ H ₆	15.6	80			0.97 ± 0.10	0.193 ± 0.018
H ₂ CO	14.6	100			1.69 ± 0.27	0.140 ± 0.021
H ₂ CO	15.6	100			0.75 ± 0.23	0.15 ± 0.05
NH ₃	9.5	100			<18 (3σ)	<0.52 (3σ)
NH ₃	14.5	87			<1.9 (3σ)	<0.16 (3σ)
NH ₃	15.6	73			<1.7 (3σ)	<0.34 (3σ)
C ₂ H ₂	9.5	100			<3.5 (3σ)	<0.10 (3σ)
C ₂ H ₂	14.6	87			0.31 ± 0.13‡	0.026 ± 0.011
C ₂ H ₂	15.6	73			<0.30 (3σ)	<0.060 (3σ)
73P/Schwassmann–Wachmann 3-C						
H ₂ O	3.5	100	3.0	>50	817 ± 98	
H ₂ O	9.6	100	3.0	>50	1,136 ± 154	
H ₂ O	14.5	92 ⁺⁴ ₋₃	3.2 ± 0.4	>37	712 ± 48	
H ₂ O	15.6	107 ⁺⁵ ₋₄	3.0 ± 0.2	>37	1,090 ± 62	
CO	3.6	80			<21 (3σ)	<2.6 (3σ)
CH ₃ OH	14.5	80			1.06 ± 0.21	0.149 ± 0.029
CH ₃ OH	15.6	90			2.77 ± 0.41	0.254 ± 0.037
HCN	9.6	90			3.4 ± 0.9	0.30 ± 0.09
HCN	14.5	77 ± 2			1.72 ± 0.17	0.242 ± 0.014
HCN	15.6	88 ± 3			2.30 ± 0.22	0.211 ± 0.010
C ₂ H ₆	3.5	80			0.53 ± 0.26‡	0.065 ± 0.033
C ₂ H ₆	9.6	90			1.3 ± 0.4	0.11 ± 0.03
C ₂ H ₆	14.5	88 ⁺²⁴ ₋₁₆			0.77 ± 0.08	0.107 ± 0.011
C ₂ H ₆	15.6	100			1.30 ± 0.13	0.119 ± 0.011
H ₂ CO	14.5	100			1.05 ± 0.25	0.147 ± 0.033
H ₂ CO	15.6	100			1.04 ± 0.22	0.095 ± 0.020
NH ₃	14.5	77			<2.4 (3σ)	<0.33 (3σ)
NH ₃	15.6	88			<3.2 (3σ)	<0.29 (3σ)
C ₂ H ₂	9.6	90			<5.0 (3σ)	<0.44 (3σ)
C ₂ H ₂	14.5	77			0.35 ± 0.14‡	0.049 ± 0.020
C ₂ H ₂	15.6	88			<0.36 (3σ)	<0.033 (3σ)

* Calculated rotational temperatures (T_{rot}), ortho-to-para ratios (OPRs), production rates and abundances are listed with error values (± 1 s.e.m.), which incorporate photon noise, calibration uncertainties, and line-by-line deviations between the data and fluorescence models. Values for absolute production rates are sensitive to model assumptions and observing circumstances that are difficult to quantify, including deviations from the spherically symmetric, uniform outflow velocity coma model, errors in the assumed outflow velocity, calibration uncertainties within a setting, uncertainties due to slit losses, temporal variability, and the possible release of volatiles outside our aperture. Therefore, the error bars on absolute abundances may be underestimated. Relative abundances are less sensitive to these factors because most of these species are detected simultaneously within the same grating setting as H₂O (Supplementary Table S1), thereby eliminating most sources of systematic error. T_{rot} and OPR entries with no error bars are adopted values. Rotational temperatures and OPRs were determined within a 3 spectral × 9 spatial pixel ($0.43'' \times 1.74''$) extract centred on the nucleus.

† Nuclear spin temperatures (T_{spin}) for water are based on calculated OPRs.

‡ Production rates and relative abundances were determined on the basis of positive flux of less than 3σ using available unblended lines. A positive flux was measured over the two available and apparently unblended ν_3 C₂H₂ lines in both fragments on 14 May. Additional C₂H₂ lines are sampled in this spectral region, but many fall in regions of poor atmospheric transmittance or are probably blended with other species. In addition, coadding weak lines in a spectral region with many emissions (Fig. 1) increases the chance for contamination from unknown or unaccounted for species. Therefore, these 'detections' are not considered firm and are regarded as upper limits. Sample line fluxes and g -factors are given in Supplementary Table S2.

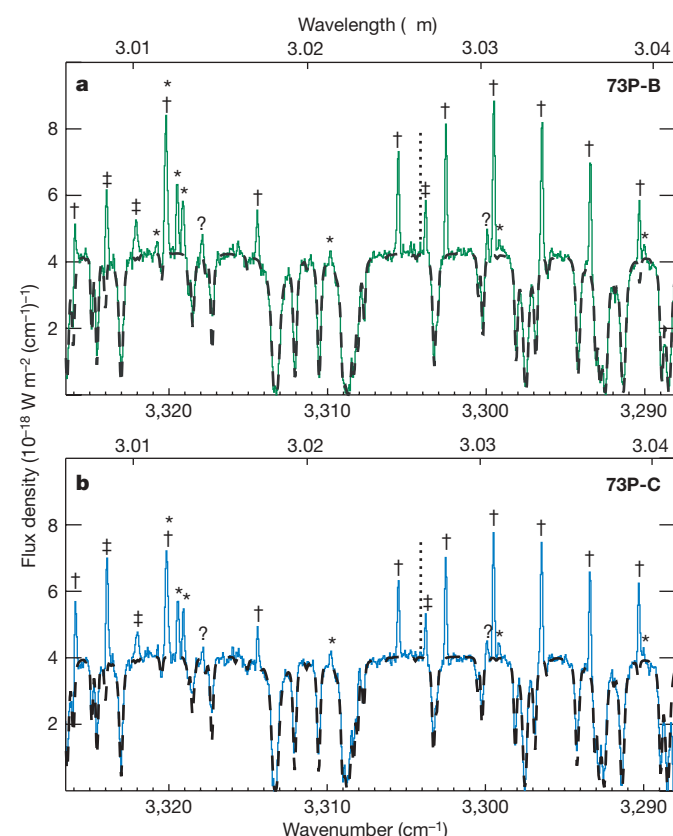


Figure 1 | NIRSPEC spectra of the B and C fragments of comet 73P/Schwassmann–Wachmann 3. Selected high-resolution ($\lambda/\Delta\lambda \approx 25,000$) spectra showing the detection of several species: *H₂O, †HCN, ‡OH, ?unidentified. Spectra are co-added over 3 spectral × 9 spatial pixel extracts ($0.43'' \times 1.74''$) centred on the peak of the dust continuum. Solid green traces, 73P-B spectra; solid blue traces, 73P-C spectra; and dashed black traces, best-fit synthetic atmospheric models. Molecular lines from the comet are shown as emissions in the comet spectra above the atmospheric model. The 1σ noise level is $\sim 1 \times 10^{-19} \text{ W m}^{-2} \text{ cm}^{-1}$ in the spectra. Spectral coverage is given in both μm (top) and cm^{-1} (bottom). **a**, 73P-B on UT May 14.6, showing detections of HCN, H₂O and OH (48 min on-source). C₂H₂ emission lines are sampled but are absent in the spectrum. **b**, 73P-C on UT May 15.6, covering the same spectral region as **a** (32 min on-source). Clear detections of HCN, H₂O and OH are seen, but as in **a**, C₂H₂ emissions are absent. The dotted vertical lines near $3,304 \text{ cm}^{-1}$ denote the expected position of the strongest unblended C₂H₂ line ($\nu_3 \text{ R3}$) in both frames. Spectra of the B and C fragments are very much alike (much more similar to each other than to spectra of other sampled comets), and suggest a homogeneous volatile composition for the original intact nucleus of 73P. See also Supplementary Fig. S1.

different dynamical classes, it would not be surprising if their nuclei were structurally dissimilar (homogeneous versus heterogeneous). However, we note that previous measurements of fragmented comets lacked either the sensitivity or the simultaneity of the 73P observations, making interpretation in the same context as the 73P data difficult (Supplementary Information). The extraordinarily favourable

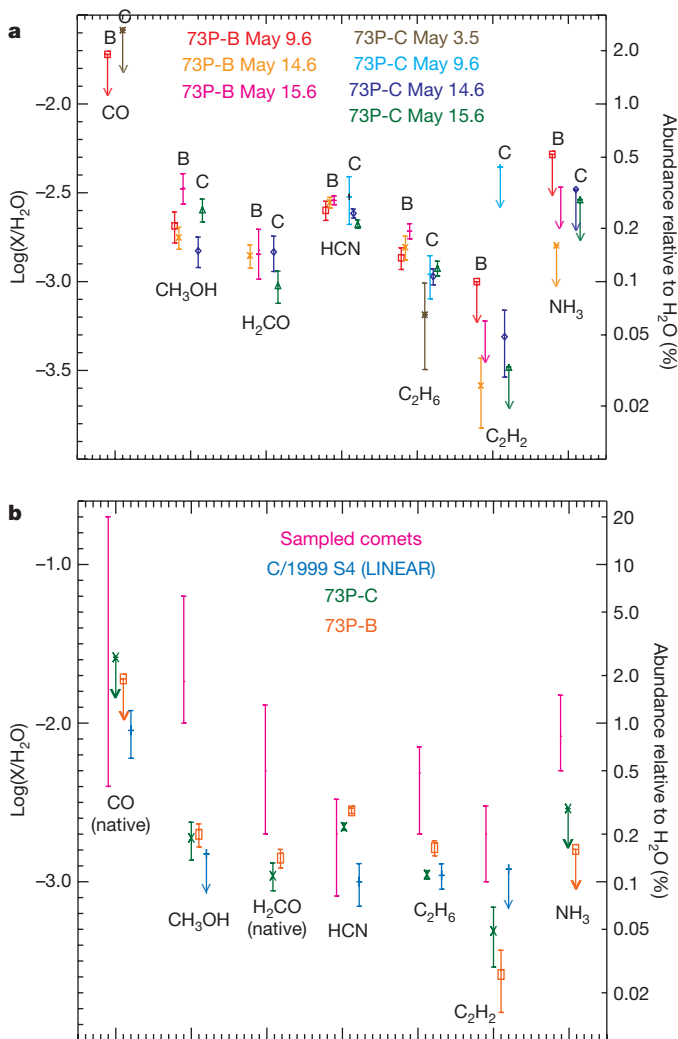


Figure 2 | Relative abundances of detected species in 73P/Schwassmann-Wachmann 3 and comparison to other comets. **a**, Relative abundances for all measured species in 73P/Schwassmann-Wachmann 3. Measurements by date in both fragments B and C are plotted with ± 1 s.e.m. error bars (see also Table 1). 3σ upper limits are given by downward arrows. This illustrates that relative abundances are similar from day to day and between fragments. **b**, A comparison of relative volatile abundances in 73P-B and 73P-C, with other comets. The range of measured abundances from the overall comet population (excepting C/1999 S4 (LINEAR)) is denoted by 'Sampled comets'^{23,3}. Abundances for C/1999 S4 (LINEAR) are also plotted³⁰. Relative abundances in 73P represent a weighted average of abundances determined on the dates measured with ± 1 s.e.m. error bars. For species that were not detected, the 3σ upper limits are given by downward arrows and represent the most constraining values. For C_2H_2 , the tentative detections from UT May 14.6 are plotted (see Supplementary Information). We note that comparisons of H_2CO (and to a lesser extent CO) in 73P to other comets are complicated by the possible presence of extended sources (see Supplementary Information). This figure illustrates that relative abundances in 73P (with the exception of HCN) resemble those in C/1999 S4 (LINEAR) and are depleted relative to the range of values observed for the general comet population. We note that 73P and C/1999 S4 (LINEAR) are compositionally similar, even though they are from different dynamical classes and probably formed in different nebular regions, further illustrating the difficulties in relating comet chemistry to nebular formation region.

2006 apparition of 73P was unique in that multiple parent volatiles were detected or constrained in two distinct fragments nearly simultaneously with high sensitivity.

Sampling the interior of the Jupiter-family comet 9P/Tempel 1 (hereafter 9P) was a primary goal of the Deep Impact mission²², so a comparison with 73P, also probably from the Kuiper belt reservoir, is desirable. Spacecraft and ground-based analyses of the pre-impact coma and the post-impact volatile ejecta may suggest a heterogeneous composition for 9P (refs 22, 23). Whether the heterogeneity in 9P is due to the accretion of chemically diverse material in the formation epoch or to evolutionary processing with depth is unclear (Supplementary Information). We note that the volatile chemistries of these comets are different; 73P is classified as carbon-chain depleted, whereas 9P is in the carbon-chain normal class (although on the depleted end of normal)¹. This suggests that 73P and 9P experienced different formative conditions or processing histories (or both), despite probably originating from the same reservoir.

Our results also show that in addition to a homogeneous composition, both 73P-B and 73P-C are depleted in all measured volatile species with respect to H_2O , except for HCN (Table 1, Fig. 2, Supplementary Information). Depletion of volatile ices in 73P could indicate that its composition has thermally evolved, even in the deep interior of its nucleus, during numerous perihelion passages; however, modelling studies suggest that thermal processing affects only the outer layers of the nucleus and not the deep interior²⁴. Thermal processing alone also cannot explain why the ratio of CH_3OH to HCN, species with similar volatilities, is about a factor of ten depleted in 73P compared to what is typically seen in comets (Fig. 2b)^{2,3}.

Kuiper belt comets with depleted volatile abundances are common, as more than half of Jupiter-family comets measured with narrowband photometry are in the depleted carbon-chain class¹. In contrast, a significantly smaller fraction (about 20%) of comets depleted in organic volatiles and carbon-chain species have been observed in the long-period Oort cloud population¹⁻³. It has been suggested that these differences are independent of dynamical age, and thus reflect primitive comet chemistry¹. If so, this suggests that Kuiper belt comets formed from source material that was more volatile-poor, on average, than material that accreted into Oort cloud comets. This is counterintuitive, because a colder formation region (the Kuiper belt for Jupiter-family comets versus the giant planets' region for Oort cloud comets) should lead to a higher concentration of the most volatile species in the ices that accreted to form Kuiper belt comets. The distinction between different nebular formation regions, however, is blurred by significant turbulent mixing of materials from the inner and outer solar nebula²⁵⁻²⁷, as well as by the migration of the giant planets²⁸. If the nebula lifetime was long compared to the time required to accrete comets, the timing of comet formation may rival (or even surpass) the importance of formation region in determining the chemical inventory and structure of cometary nuclei²⁹.

The measured composition of parent volatiles from the deep interior of 73P is consistent with the idea that the depleted carbon-chain chemistry measured in at least some Jupiter-family comets is primordial. However, results suggesting a heterogeneous composition for 9P may indicate that this is not the case for all Jupiter-family comets, and illustrate the difficulty of relating measured comet chemistry to nebular formation region. Whether the timing of comet formation or another as yet unidentified effect primarily causes these differences remains to be seen.

Received 7 March; accepted 2 May 2007.

1. A'Hearn, M. F., Millis, R. L., Schleicher, D. G., Osip, D. J. & Birch, P. V. The ensemble properties of comets: Results from narrowband photometry of 85 comets, 1976–1992. *Icarus* **118**, 223–270 (1995).
2. Bockelée-Morvan, D., Crovisier, J., Mumma, M. J. & Weaver, H. A. in *Comets II* (eds Festou, M. C., Keller, H. U. & Weaver, H. A.) 391–423 (Univ. Arizona Press, Tucson, 2005).

3. Biver, N. *et al.* Chemical composition diversity among 24 comets observed at radio wavelengths. *Earth Moon Planets* **90**, 323–333 (2002).
4. Stern, S. A. The evolution of comets in the Oort cloud and Kuiper belt. *Nature* **424**, 639–642 (2003).
5. A'Hearn, M. F., Thurber, C. H. & Millis, R. L. Evaporation of ices from Comet West. *Astron. J.* **82**, 518–524 (1977).
6. Bockelée-Morvan, D. *et al.* Outgassing behavior and composition of comet C/1999 S4 (LINEAR) during its disruption. *Science* **292**, 1339–1343 (2001).
7. Gibb, E. L. *et al.* The organic composition of C/2001 A2 (LINEAR): II. Search for heterogeneity within a comet nucleus. *Icarus* **188**, 224–232 (2007).
8. Villanueva, G. *et al.* The volatile composition of the split ecliptic comet 73P/Schwassmann-Wachmann 3: A comparison of fragments C and B. *Astrophys. J.* **650**, L87–L90 (2006).
9. Fink, U. & Hicks, M. D. A survey of 39 comets using CCD spectroscopy. *Astrophys. J.* **459**, 729–743 (1996).
10. Schleicher, D., Birch, P. V. & Bair, A. N. The composition of the interior of comet 73P/Schwassmann-Wachmann 3: Results from narrowband photometry of multiple components. *Bull. Am. Astron. Soc.* **38**, 485 (2006).
11. Boehnhardt, H. *et al.* 73P/Schwassmann-Wachmann 3 – one orbit after breakup: Search for fragments. *Earth Moon Planets* **90**, 131–139 (2002).
12. Crovisier, J. *et al.* What happened to comet 73P/Schwassmann-Wachmann 3? *Astron. Astrophys.* **310**, L17–L20 (1996).
13. Weaver, H. A. *et al.* Hubble Space Telescope investigation of the disintegration of 73P/Schwassmann-Wachmann 3. *Bull. Am. Astron. Soc.* **38**, 490 (2006).
14. Nolan, M. C. *et al.* Radar observations of comet 73P/Schwassmann-Wachmann 3. *Bull. Am. Astron. Soc.* **38**, 504 (2006).
15. Greene, T. P., Tokunaga, A. T., Toomey, D. W. & Carr, J. B. CSHELL: A high spectral resolution 1–5 μm cryogenic echelle spectrograph for the IRTF. *Proc. SPIE* **1946**, 313–324 (1993).
16. McLean, I. S. *et al.* The design and development of NIRSPEC: A near-infrared echelle spectrograph for the Keck II telescope. *Proc. SPIE* **3354**, 566–578 (1998).
17. Dello Russo, N., DiSanti, M. A., Mumma, M. J., Magee-Sauer, K. & Rettig, T. W. Carbonyl sulfide in comets C/1996 B2 (Hyakutake) and C/1995 O1 (Hale-Bopp): Evidence for an extended source in Hale-Bopp. *Icarus* **135**, 377–388 (1998).
18. Mumma, M. J. *et al.* A survey of organic volatile species in comet C/1999 H1 (Lee) using NIRSPEC at the Keck Observatory. *Astrophys. J.* **546**, 1183–1193 (2001).
19. Dello Russo, N. *et al.* Water production and release in comet 153P/Ikeya-Zhang (C/2002 C1): Accurate rotational temperature retrievals from hot-band lines near 2.9- μm . *Icarus* **168**, 186–200 (2004).
20. Reuter, D. C., Mumma, M. J. & Nadler, S. Infrared fluorescence efficiencies for the ν_1 and ν_5 bands of formaldehyde in the solar radiation field. *Astrophys. J.* **341**, 1045–1058 (1989).
21. Dello Russo, N., Mumma, M. J., DiSanti, M. A., Magee-Sauer, K. & Novak, R. Ethane production and release in comet C/1995 O1 Hale-Bopp. *Icarus* **153**, 162–179 (2001).
22. A'Hearn, M. F. *et al.* Deep Impact: Excavating comet Tempel 1. *Science* **310**, 258–264 (2005).
23. Mumma, M. J. *et al.* Parent volatiles in comet 9P/Tempel 1: Before and after impact. *Science* **310**, 270–274 (2005).
24. Prialnik, D., Benkhoff, J. & Podolak, M. in *Comets II* (eds Festou, M. C., Keller, H. U. & Weaver, H. A.) 359–387 (Univ. Arizona Press, Tucson, 2005).
25. Morfill, G. E. Some cosmochemical consequences of a turbulent protoplanetary cloud. *Icarus* **53**, 41–54 (1983).
26. Bockelée-Morvan, D., Gautier, D., Hersant, F., Huré, J.-M. & Robert, F. Turbulent radial mixing in the solar nebula as the source of crystalline silicates in comets. *Astron. Astrophys.* **384**, 1107–1118 (2002).
27. Lisse, C. M. *et al.* Spitzer spectral observations of the Deep Impact ejecta. *Science* **313**, 635–640 (2006).
28. Levison, H. F. & Morbidelli, A. The formation of the Kuiper belt by the outward transport of bodies during Neptune's migration. *Nature* **426**, 419–421 (2003).
29. Nuth, J. A., Hill, H. G. M. & Kletetschka, G. Determining the ages of comets from the fraction of crystalline dust. *Nature* **406**, 275–276 (2000).
30. Mumma, M. J. *et al.* The startling organic composition of C/1999 S4 (LINEAR): A comet formed near Jupiter? *Science* **292**, 1334–1339 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The NASA Planetary Atmospheres and Planetary Astronomy Programs supported this work. Data were obtained at the NASA IRTF operated by the University of Hawaii under cooperative agreement with the NASA Science Mission Directorate, Planetary Astronomy Program. Data were also obtained at the W. M. Keck Observatory, which is operated as a scientific partnership among the California Institute of Technology, the University of California and NASA. The Observatory was made possible by the financial support of the W. M. Keck Foundation. We acknowledge the assistance and expertise of B. Golisch, P. Sears, D. Griep, G. Hill and G. Puniwai, who helped to successfully obtain these data under difficult conditions; and we acknowledge that the summit of Mauna Kea, from which we had the privilege of conducting these observations, is revered within the indigenous Hawaiian community.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to N.D.R. (neil.dello.russo@jhuapl.edu).

LETTERS

High-resolution spectroscopy of two-dimensional electron systems

O. E. Dial¹, R. C. Ashoori¹, L. N. Pfeiffer² & K. W. West²

Spectroscopic methods involving the sudden injection or ejection of electrons in materials are a powerful probe of electronic structure and interactions¹. These techniques, such as photoemission and tunnelling, yield measurements of the 'single-particle' density of states spectrum of a system². This density of states is proportional to the probability of successfully injecting or ejecting an electron in these experiments. It is equal to the number of electronic states in the system able to accept an injected electron as a function of its energy, and is among the most fundamental and directly calculable quantities in theories of highly interacting systems³. However, the two-dimensional electron system (2DES), host to remarkable correlated electron states such as the fractional quantum Hall effect⁴, has proved difficult to probe spectroscopically. Here we present an improved version of time-domain capacitance spectroscopy⁵ that allows us to measure the single-particle density of states of a 2DES with unprecedented fidelity and resolution. Using the method, we perform measurements of a cold 2DES, providing direct measurements of interesting correlated electronic effects at energies that are difficult to reach with other techniques; these effects include the single-particle exchange-enhanced spin gap⁶, single-particle lifetimes⁷ in the quantum Hall system, and exchange splitting of Landau levels not at the Fermi surface.

Spectroscopy of the 2DES has been performed optically and using electron tunnelling. Traditional optical methods probe long-wavelength collective excitations rather than the single-particle density of states (SPDOS)⁸. Attempts to measure the SPDOS optically have relied on the introduction of local impurities, altering the measured spectrum⁹. Approaches based on tunnelling spectroscopy have been hampered by several difficulties^{10–12}. Typically, the tunnel current enters the 2DES perpendicularly to the plane through a tunnel barrier. In steady-state measurements, maintenance of this current requires it to have a path to exit the 2DES. The current can either flow in the plane of the 2DES to a distant contact, or it can tunnel out perpendicularly from the other side of the plane. However, if the in-plane conductivity of the device is small (as frequently occurs in quantum Hall systems), the current cannot flow out through the plane of the 2DES. If the current tunnels out perpendicularly through a second barrier, the two barriers form a resistive divider and variations in the tunnel current alter the 2D density and tunnel voltage. Finally, ohmic heating limits the useful range of both these techniques to small tunnel voltages.

Time-domain capacitance spectroscopy (TDCS) eliminates all of the major problems associated with 2D tunnelling spectroscopy by abandoning steady-state measurement. In TDCS, we replace the ohmic contact to the 2DES with a distant, electrically isolated electrode parallel to the plane of the 2DES. We capacitively apply the tunnel voltage using the isolated electrode, and then capacitively detect the tunnelled electrons rather than removing them through

a second contact or barrier. Because the electron does not need to move in the plane of the 2DES to be detected, we can measure the SPDOS even when the 2DES is insulating or empty. Although this capacitive coupling requires us to use a pulsed a.c. measurement to extract current–voltage (I – V) characteristics, it also allows us to use very short duty cycles (typically 0.01%) to minimize heating.

In our measurements, the 2DES is sandwiched between two electrodes of a 'tunnel capacitor' created from a GaAs heterostructure (Fig. 1a, b). The tunnelling electrode, used as a source for injected electrons, is separated from the 2DES by a thin $\text{Al}_{0.33}\text{Ga}_{0.67}\text{As}$ tunnelling barrier. The isolated electrode, used to apply voltages and detect charges, is separated from the 2DES by a thicker, insulating barrier. To minimize scattering, there is no intentional doping in either barrier.

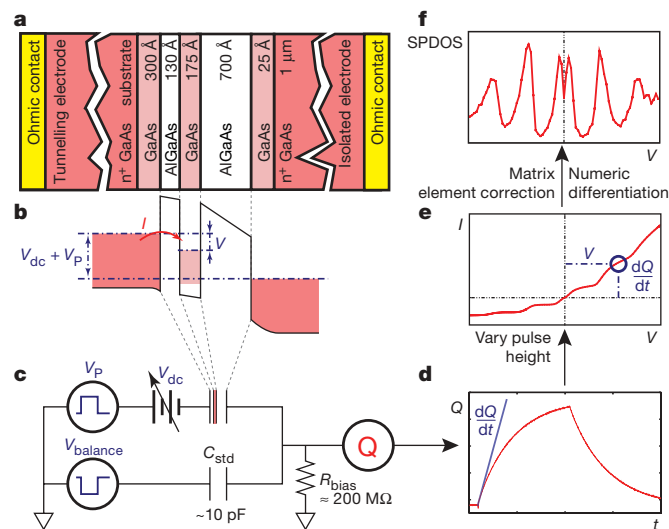


Figure 1 | Time-domain capacitance spectroscopy. **a**, GaAs heterostructure used to construct the tunnel capacitor. The only ohmic contacts are to the top and bottom three-dimensional doped regions, used as electrodes. **b**, Band structure of the tunnel capacitor with an applied d.c. bias V_{DC} in the instant after a pulse of height V_P is applied across it, showing the well defined voltage V between the tunnelling electrode and 2DES, and the tunnel current (I). **c**, Simplified diagram of circuitry used to perform the measurement, including a standard capacitor C_{std} used with a pulse of opposite polarity $V_{balance}$ to remove the large charge step from the background geometric capacitance of the structure and a resistor R_{bias} used to maintain the d.c. voltage across the sample. **d**, Measured charge traces, taken with a long pulse for illustrative purposes, showing the charge response averaged over 250,000 repetitions. The slope of this curve immediately after the pulse is applied (blue line) gives the tunnel current. **e**, I – V characteristic for tunnelling into the 2DES constructed by combining data from ~ 200 different applied pulse heights. **f**, SPDOS spectrum attained by multiplying **e** by a smooth matrix element correction and differentiating.

¹Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. ²Bell Laboratories, Alcatel-Lucent, Murray Hill, New Jersey 07974, USA.

The 2D electron density is controlled from 0 up to $4 \times 10^{11} \text{ cm}^{-2}$ by varying the d.c. voltage applied across the electrodes.

To perform our measurement, we apply a sudden voltage step V_p (≤ 1 ns rise time), disequilibrating the tunnelling electrode from the 2DES. In the absence of bridge compensation, charging the geometric capacitance (~ 10 pF) between the electrodes would generate a large charge step at the detector. We remove this step by applying a pulse of opposite polarity to a standard capacitor C_{std} connected to the isolated electrode (Fig. 1c). As electrons tunnel across the barrier, we measure a long slow rise in the total charge on the isolated electrode due to their image charge. Because electrons tunnel only a short distance into the tunnel capacitor to reach the 2DES, this charging signal is much smaller ($<33\%$) than the uncompensated initial charge step. The charging signal is measured using a cryogenic amplifier and massive ($250,000\times$) repetitive signal averaging. We fit the derivative of this signal as a function of time to determine the tunnelling current (Fig. 1d). In the first instant after the step, the voltage V between the 2DES and the tunnelling electrode is precisely V_p multiplied by an easily calibrated geometric lever-arm (defined as the ratio of the distance between the 2DES and the tunnelling electrode to the distance between the isolated electrode and the tunnelling electrode). This, together with the measured tunnelling current, gives us a single point on an I - V curve; varying the pulse amplitude traces out the entire curve (Fig. 1e).

With continuous application of the pulsed voltage, the 2DES and electrode would re-equilibrate in a few milliseconds. However, we use only the first ~ 100 ns of the step to determine the current, before the voltage across the tunnel barrier has relaxed significantly. This short time interval means that little charge is transferred, so electrons are tunnelling into and out of cold electronic systems. After this brief measurement window has elapsed, the voltage step is removed in preparation for the next pulse, and the system is allowed to re-equilibrate.

Improving on our prior work⁵, with precise control over the shape of the applied pulses, we are now able to reduce the systematic and random errors in this I - V curve enough to numerically differentiate it (Fig. 1f), giving dI/dV , a quantity proportional to the SPDOS multiplied by a slowly varying tunnelling matrix element (see Supplementary Information). Although the energy resolution is fundamentally limited only by temperature (see below), the colour-scale plots are taken with lower resolution to reduce acquisition time.

Figure 2 shows typical plots of the SPDOS as a function of energy and Landau level filling fraction ν . We extract the value of the geometric lever-arm by measuring the known cyclotron splitting of the empty well. We estimate the maximum error that accrues in the energy calibration (the error in the y axis) as we move away from zero density and energy to be about 1%, predominantly due to variation of the lever-arm from polarization of the 2DES and depletion of the electrodes. From left to right within each plot, the quantum well starts fully depleted, with Landau levels visible above the Fermi energy ($E > 0$). As the well begins to fill, a magnetic-field-induced Coulomb gap^{5,13–15}, appearing as a dark band across $E = 0$ in each plot, opens at the Fermi surface. Each Landau level is pinned by its high density of states as it crosses the Fermi energy. This results in a stair-step pattern of successive flat pinned regions followed by downward steps corresponding to chemical potential jumps as the next Landau level is pulled down to the Fermi energy¹⁶.

Exchange-enhanced spin-splitting is visible as splitting of the Landau levels as they cross the Fermi surface. At even-integer filling, there is no net spin polarization, causing the exchange gap to vanish. As the density is raised, one spin state of the orbital Landau level begins to fill, and the attractive exchange interaction creates a gap to the other spin state of the orbital level that grows until it saturates with one completely filled spin state at odd-integer ν . The gap shrinks again as the upper spin state is filled and pulled down in energy to meet the lower spin state. This gives rise to the 'oscillatory g-factor' in quantum Hall systems⁶ and, together with the pinning of the spin

state being filled to the Fermi energy, creates a distinctive pattern as each Landau level crosses the Fermi surface.

We measure the spin-splittings E_j at odd-integer filling by fitting the peaks in constant ν line-cuts of the spectra to Lorentzians, and develop an empirical formula for describing the splittings. The results are shown in Fig. 3a. Varying density at fixed values of the magnetic field B , we find that the spin gaps diminish with a $1/\sqrt{\nu}$ dependence (Fig. 3b). To compare the B dependence of splittings at different ν , Fig. 3c shows the splittings multiplied by $\sqrt{\nu}$ as a function of B . The data cleanly collapse onto a single line, revealing a linear dependence on B . The splittings are well characterized by $E_j = 1.2 \pm 0.03 B/\sqrt{\nu}$, where we have forced the fit through the origin.

No previous experiment has directly measured the single-particle exchange gap that arises in the sudden introduction of a single electron. However, an exchange-enhanced gap is also present in the 'thermodynamic' density of states often measured using capacitance measurements. These measurements monitor the change in the 2DES chemical potential long after adding electrons to the system, allowing time for correlations to form^{17,18}.

Nonetheless, a linear dependence on B has also been seen both in capacitance measurements of the gap in the thermodynamic density of states¹⁸ and in thermally activated measurements of the transport gap¹⁹. However, the measured values of E_j are roughly a factor of 4 or 2 smaller than our measurements, respectively. In contrast with our results, thermal activation experiments find no ν dependence at fixed field^{19,20}.

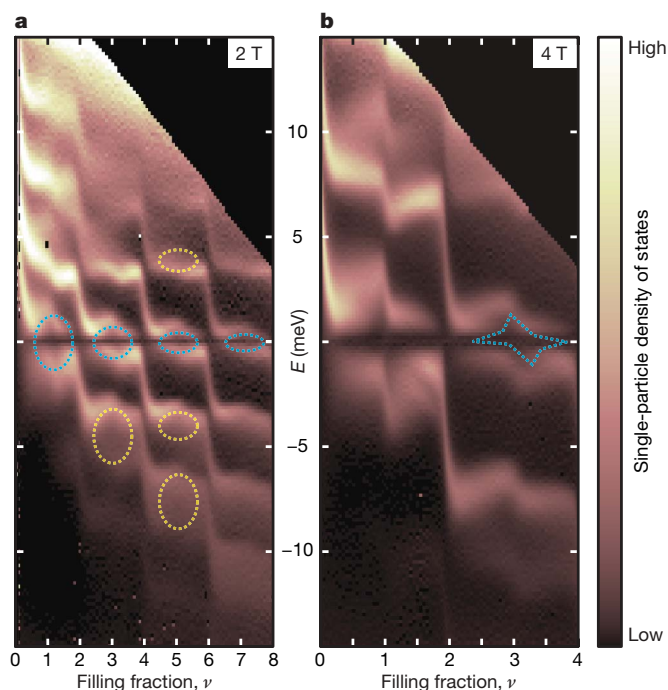


Figure 2 | Single-particle density of states spectra. The spectra are taken at fixed magnetic fields of 2 T (a) and 4 T (b), with an energy resolution of 170 μeV and taken at approximately 100 mK. Each bright peak corresponds to a Landau level. The vertical (y) axis is the tunnelling energy measured with respect to the Fermi energy. Energies greater than zero correspond to injecting electrons into empty states in the quantum well, whereas energies less than zero correspond to ejecting electrons from filled states. A smooth approximate correction for variation of the tunnelling matrix element has been applied (see Supplementary Information), resulting in a dark band in the top-right corner of each plot. The horizontal (x) axis is filling fraction ν (proportional to electron density at fixed magnetic field). The applied d.c. gate voltage is transformed to ν through a.c. capacitance measurements¹⁷. Dashed blue ellipses in a mark exchange-enhanced spin splittings, while dashed yellow ellipses mark some of the indirect exchange splittings. The blue dashed shape in b shows the expected behaviour of the Landau levels as they cross the Fermi energy.

In the absence of Landau level mixing, the exchange-enhanced spin-gap at the Fermi energy, E_j , is calculated^{21,22} to grow as \sqrt{B} at fixed ν and to be largest at $\nu = 1$ compared with other odd-integer fillings. However, we expect that mixing is important; for $B < 6.5$ T, the exchange energy scale $e^2/\epsilon l_b$ is greater than the cyclotron splitting $\hbar\omega_c$, and level repulsion from nearby Landau levels should limit the exchange splitting to approximately $\hbar\omega_c$. Here, e is the electron charge, ϵ is the dielectric constant, l_b is the magnetic length $\sqrt{\hbar/eB}$, ω_c is the cyclotron frequency eB/m^* , and m^* is the electron effective mass. The linear dependence of E_j on B has been found in structure-specific calculations²³. More recent theory, providing results only at $\nu = 1$, predicts a splitting, linear with B , roughly 30% smaller than our value²⁴. No theory predicts the observed $B/\sqrt{\nu}$ dependence.

We also observe the indirect exchange splitting of Landau levels above and below the Fermi surface by spin-polarized Landau levels at the Fermi surface (dashed yellow ovals in Fig. 2a). Calculations of the exchange matrix element²¹ between different orbital Landau levels indicate that because of smaller wavefunction overlap, at any given density this indirect splitting should be smaller than the exchange-enhanced spin gap at the Fermi energy (neglecting Landau level mixing). Although we find that this is sometimes correct, the most deeply buried orbital Landau level, $N = 0$, is consistently split more than any other level in the spectrum. This level is unique in that there is no deeper Landau level available for level mixing, so there is no level repulsion to limit the exchange splitting on the lower energy side. We suggest that level mixing generally reduces these indirect exchange splittings just as it reduces the splitting at the Fermi surface, but in the case of $N = 0$ the absence of lower energy Landau levels allows the splitting to grow larger.

The most remarkable aspect of this observation is that the split Landau levels are at energies several hundred times the thermal energy, yet their splitting reflects the relatively delicate exchange splitting at the Fermi surface. This demonstrates that the high energy features in the spectra depend on properties of the many-body system only present at low temperatures. The high energy spectra constitute a previously unexplored probe of the highly correlated many-body ground state.

The lineshapes of Landau levels away from the Fermi energy provide a window into the effects of electron–electron interactions. We find that the lineshape at fields of 1 T or less fits well to lorentzians

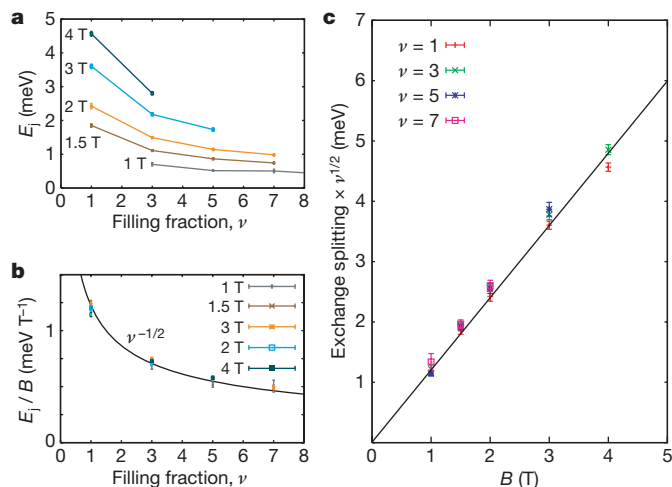


Figure 3 | Exchange-enhanced spin splitting at the Fermi energy. **a**, The unscaled exchange-enhanced spin gap as a function of odd-integer ν and field. **b**, $\nu^{-1/2}$ dependence (solid curve) shown by scaling the splitting as a function of ν by $1/B$. **c**, E_j scaled by $\sqrt{\nu}$, showing the proportionality of E_j to $B/\sqrt{\nu}$. The Landau levels are fitted by a superposition of lorentzians to extract the splittings. Error bars show the standard deviation of the exchange splitting, as determined by the variation of the fit residuals.

(Fig. 4a, bottom left). With a nearly empty quantum well, sharp Landau levels are observed at any energy (blue box in Fig. 4a). As the 2DES is filled, the levels far from the Fermi surface suddenly broaden and disappear. In an occupied 2DES, the injected electron can scatter off electrons already present in the Fermi sea. The phase space available for this scattering grows rapidly as the injection energy moves away from the Fermi energy. At large energies, the short lifetime due to electron–electron scattering becomes the dominant source of broadening. By comparison, at low energies we find maxima in the peak width at even-integer ν and minima at odd-integer ν . Here, disorder is the chief source of broadening. As observed, this disorder-induced broadening should display maxima at even-integer ν , where the density of states of the 2DES at the Fermi level is small and the 2DES screens poorly. It shows minima elsewhere, where the density of states is high and the 2DES screens disorder well.

In the case of a disorder-free 2DES, the electron lifetime τ_{ee} has been calculated at zero field by Chaplik⁷ (see Supplementary Information). At small magnetic fields such that the cyclotron energy is small compared to the Coulomb interaction energy, we expect this result to hold. Fitting the widths of the Landau levels to lorentzians gives us an irregularly spaced mesh of broadenings as a function of energy and density. In order to simplify comparisons, we fit a smooth function through the observed width of the Landau levels at each density, and plot contours of constant broadening (Fig. 4b). At high energies and away from even-integer ν , we see good agreement with Chaplik's result over a broad range of densities with no adjustable

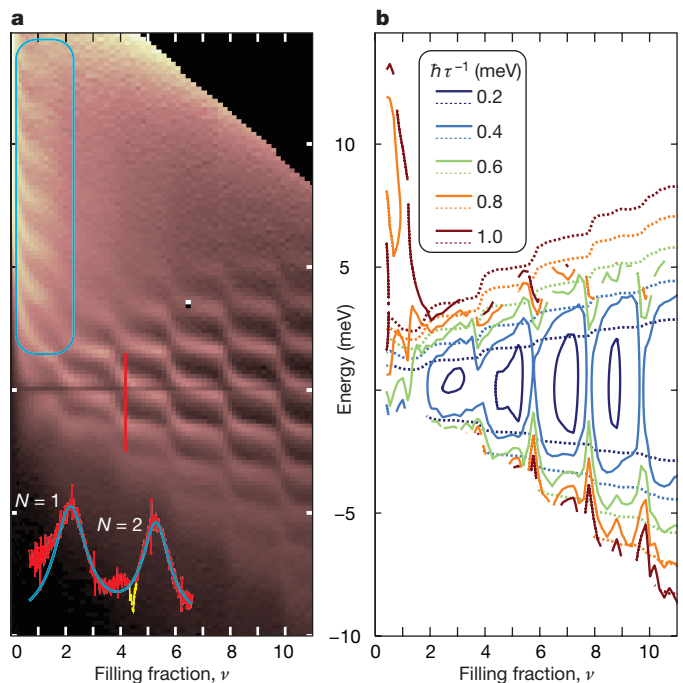


Figure 4 | Quasiparticle lifetimes. **a**, TDCS spectra at 1 T. At low densities all Landau levels in the spectrum are resolved (blue box at top left), while at higher densities the short lifetime due to electron–electron interactions broadens the Landau levels farthest from the Fermi surface until they are not resolved. A line cut taken through the red line in the main plot (approximate centre) with temperature-limited resolution and fitted by a superposition of two lorentzians is shown superposed at bottom left. The Coulomb gap is highlighted in yellow in this line cut. Deviations of the fit at large negative energies (left) are from the tail of the $N = 0$ Landau level, not included in the fit. **b**, Solid lines, contours of constant Landau level half-width at half-maximum, as obtained by fitting the peaks in **a** with lorentzians and interpolating between peaks with a fourth-order polynomial. We equate this level width Γ with a lifetime τ using $\Gamma = \hbar/\tau$. Dashed lines, \hbar/τ_{ee} from Chaplik's result⁷ with no adjustable parameters or scaling (see Supplementary Information).

parameters; this is the region where disorder broadening is small and we expect lifetime broadening to dominate. Prior measurements of momentum and phase scattering times have found quantitative agreement with Chaplik's result using an adjustable scaling parameter, but these data represent the first direct measurement of the single-particle lifetime^{25,26}.

Received 26 December 2006; accepted 4 January 2007.

1. Giaever, I. Energy gap in superconductors measured by electron tunneling. *Phys. Rev. Lett.* **5**, 147–148 (1960).
2. Bardeen, J. Tunneling from a many-particle point of view. *Phys. Rev. Lett.* **6**, 57–59 (1961).
3. Hedin, L., Lundqvist, B. I. & Lundqvist, S. in *Electronic Density of States* (ed. Bennett, L. H.) 233–249 (NBS Spec. Publ. 323, National Bureau of Standards, Washington DC, 1971).
4. Stormer, H. L., Tsui, D. C. & Gossard, A. C. The fractional quantum Hall effect. *Rev. Mod. Phys.* **71**, S298–S305 (1999).
5. Chan, H. B., Glicofridis, P. I., Ashoori, R. C. & Melloch, M. R. Universal linear density of states for tunneling into the two-dimensional electron gas in a magnetic field. *Phys. Rev. Lett.* **79**, 2867–2870 (1997).
6. Ando, T. & Uemura, Y. Theory of oscillatory g factor in an MOS inversion layer under strong magnetic fields. *J. Phys. Soc. Jpn* **37**, 1044–1052 (1974).
7. Chaplik, A. V. Energy spectrum and electron scattering processes in inversion layers. *Sov. Phys. JETP* **33**, 997–1000 (1971).
8. Pinczuk, A. *et al.* Observation of roton density of states in two-dimensional Landau-level excitations. *Phys. Rev. Lett.* **61**, 2701–2704 (1988).
9. Kukushkin, I. V. & Timofeev, V. B. Magneto-optics of strongly correlated two-dimensional electrons in single heterojunctions. *Adv. Phys.* **45**, 147–242 (1996).
10. Mendez, E. E., Esaki, L. & Wang, W. I. Resonant magnetotunneling in GaAlAs-GaAs-GaAlAs heterostructures. *Phys. Rev. B* **33**, 2893–2896 (1986).
11. Eisenstein, J. P., Gramila, T. J., Pfeiffer, L. N. & West, K. W. Probing a two-dimensional Fermi surface by tunneling. *Phys. Rev. B* **44**, 6511–6514 (1991).
12. Main, P. C. *et al.* Landau-level spectroscopy of a two-dimensional electron system by tunneling through a quantum dot. *Phys. Rev. Lett.* **84**, 729–732 (2000).
13. Aleiner, I. L., Baranger, H. U. & Glazman, L. I. Tunneling into a two-dimensional electron liquid in a weak magnetic field. *Phys. Rev. Lett.* **74**, 3435–3438 (1995).
14. Ashoori, R. C., Lebens, J. A., Bigelow, N. P. & Silsbee, R. H. Equilibrium tunneling from the two-dimensional electron gas in GaAs — evidence for a magnetic-field-induced energy gap. *Phys. Rev. Lett.* **64**, 681–684 (1990).
15. Eisenstein, J. P., Pfeiffer, L. N. & West, K. W. Coulomb barrier to tunneling between parallel two-dimensional electron systems. *Phys. Rev. Lett.* **69**, 3804–3807 (1992).
16. Popov, V. G. Self-consistent equilibrium of a two-dimensional electron system with a reservoir in a quantizing magnetic field: Analytical approach. *Phys. Rev. B* **73**, 125310 (2006).
17. Ashoori, R. *The Density of States in the Two-Dimensional Electron Gas and Quantum Dots*. PhD thesis, Cornell Univ. (1991); (<http://arXiv.org/abs/cond-mat/0607739>).
18. Dolgoplov, V. T. *et al.* Direct measurements of the spin gap in the two-dimensional electron gas of AlGaAs-GaAs heterojunctions. *Phys. Rev. Lett.* **79**, 729–732 (1997).
19. Usher, A., Nicholas, R. J., Harris, J. J. & Foxon, C. T. Observation of magnetic excitons and spin waves in activation studies of a two-dimensional electron gas. *Phys. Rev. B* **41**, 1129–1134 (1990).
20. Schmeller, A., Eisenstein, J. P., Pfeiffer, L. N. & West, K. W. Evidence for skyrmions and single spin flips in the integer quantized Hall effect. *Phys. Rev. Lett.* **75**, 4290–4293 (1995).
21. MacDonald, A. H., Oji, H. C. A. & Liu, K. L. Thermodynamic properties of an interacting two-dimensional electron gas in a strong magnetic field. *Phys. Rev. B* **34**, 2681–2689 (1986).
22. Aleiner, I. L. & Glazman, L. I. Two-dimensional electron liquid in a weak magnetic field. *Phys. Rev. B* **52**, 11296–11312 (1995).
23. Smith, A. P., MacDonald, A. H. & Gumbs, G. Quasiparticle effective mass and enhanced g factor for a two-dimensional electron gas at intermediate magnetic fields. *Phys. Rev. B* **45**, 8829–8832 (1992).
24. Iordanski, S. V. & Kasbuba, A. Excitations in a quantum Hall ferromagnet with strong Coulomb interaction. *JETP Lett.* **75**, 348–353 (2002).
25. Murphy, S. Q., Eisenstein, J. P., Pfeiffer, L. N. & West, K. W. Lifetime of two-dimensional electrons measured by tunneling spectroscopy. *Phys. Rev. B* **52**, 14825–14828 (1995).
26. Yacoby, A., Sivan, U., Umbach, C. P. & Hong, J. M. Interference and dephasing by electron-electron interaction on length scales shorter than the elastic mean free path. *Phys. Rev. Lett.* **66**, 1938–1941 (1991).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by the Office of Naval Research and by the National Science Foundation funded through the NSEC Program and the MRSEC Program.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to R.C.A. (ashoori@mit.edu) or O.E.D. (dial@alum.mit.edu).

LETTERS

Reversal of the net dinitrogen gas flux in coastal marine sediments

R. W. Fulweiler¹, S. W. Nixon¹, B. A. Buckley¹ & S. L. Granger¹

The flux of nitrogen from land and atmosphere to estuaries and the coastal ocean has increased substantially in recent decades. The observed increase in nitrogen loading is caused by population growth, urbanization, expanding water and sewer infrastructure, fossil fuel combustion and synthetic fertilizer consumption^{1,2}. Most of the nitrogen is removed by denitrification in the sediments of estuaries and the continental shelf, leading to a reduction in both cultural eutrophication and nitrogen pollution of the open ocean^{3,4}. Nitrogen fixation, however, is thought to be a negligible process in sub-tidal heterotrophic marine systems⁵. Here we report sediment core data from Narragansett Bay, USA, which demonstrate that heterotrophic marine sediments can switch from being a net sink to being a net source of nitrogen. Mesocosm and core incubation experiments, together with a historic data set of mean annual chlorophyll production^{6,7}, support the idea that a climate-induced decrease in primary production has led to a decrease in organic matter deposition to the benthos and the observed reversal of the net sediment nitrogen flux. Our results suggest that some estuaries may no longer remove nitrogen from the water column. Instead, nitrogen could be exported to the continental shelf and the open ocean and could shift the effect of anthropogenic nitrogen loading beyond the immediate coastal zone.

Climate change has altered the phenology or seasonal sequencing of events in a variety of ecosystems^{8,9}. In the marine environment, the timing of phytoplankton production is particularly important because it affects both pelagic and benthic ecology^{10,11}. For shallow systems, organic matter deposited to the benthos determines fauna abundance as well as rates of biogeochemical cycling^{12,13}. In turn, the fluxes of nutrients off the sediment surface can influence water-column phytoplankton production, especially when pelagic nutrient concentrations are low. The impact of climate change on this benthic–pelagic coupling has previously been unknown.

Long-term warming^{14,15} has resulted in the loss of the winter–spring diatom bloom in most years and a decrease in mean annual chlorophyll production between the 1970s and 1990s in Narragansett Bay, Rhode Island (41.3° N, 071.1° W)⁶. This decrease has continued through 2005 and there is also a marked decrease in summer chlorophyll concentrations (Fig. 1). The exact mechanism behind this decline is not known, but it is believed to be the result of increased year-round grazing¹⁵ or an increase in cloudy days, which negatively affects the onset of the winter–spring bloom¹⁶ (Supplementary Discussion). Nitrogen inputs have remained constant for the last twenty-five years¹⁷. Combining the chlorophyll observations with associated Secchi disk data (a proxy for light penetration depth)¹⁸ in a simple empirical model that relates carbon fixation by marine phytoplankton to chlorophyll, optical depth, and surface irradiance¹⁹ suggests that primary production in the mid-bay may have declined by 40% in the last thirty years. Recent measurements in the bay have

also shown a marked decrease in sediment denitrification, sediment oxygen consumption, and in the fluxes of ammonium and phosphate from sediments to the overlying water compared with extensive measurements made in the 1970s and 1980s⁷.

As in most estuaries, denitrification has traditionally been a dominant nitrogen process in the sediments, removing between 15–25% of the nitrogen load from land and atmosphere to the bay^{3,20}. But during the summer of 2006, using the N₂/Ar technique^{21,22}, we observed a dramatic change in the net sediment N₂ flux by mid-bay sediments (Fig. 2). High rates of N-fixation were also found in triplicate cores collected in July and August at three other stations throughout the bay (Table 1). The N-fixation was seasonal, and by mid-September the sediments were once again dominated by denitrification. These N-fixation rates are remarkable for two reasons. They represent some of the highest rates measured for a marine system²³, and this is the first demonstration that heterotrophic marine sediments can switch from being a net sink to being a major net source of nitrogen. We hypothesized that the decrease in primary production led to a decrease in organic matter deposition to the benthos and the observed reversal of the net sediment N₂ flux (Supplementary Discussion). We tested this hypothesis using core incubations and nine large (4 m²) benthic mesocosms containing sediment and associated fauna from mid-bay.

The mesocosm experiment was designed to quantify the response of the benthos to different levels of organic deposition. Estimates of the proportion of phytoplankton blooms that fall to the bottom in coastal waters vary widely, with larger values for winter–spring blooms and

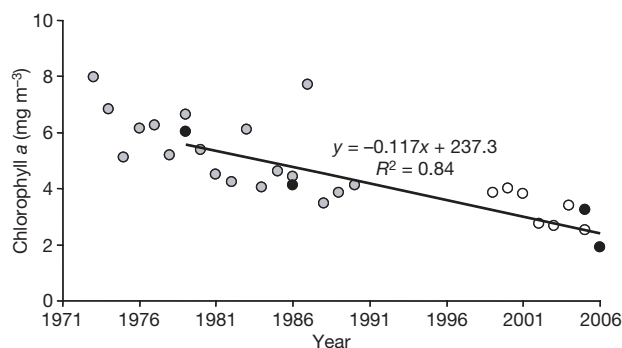


Figure 1 | Multi-decadal mean annual water column chlorophyll *a* concentrations. Mean annual water column (surface and bottom) chlorophyll *a* concentrations in mid-bay over the past three decades. Grey circles are from ref. 6; open and closed circles from the Graduate School of Oceanography plankton monitoring programme (<http://www.gso.uri.edu/phytoplankton>). Black circles are the mean summer (June, July, August) chlorophyll *a* concentrations shown for the years when N₂ fluxes were measured at the mid-bay station (41° 35.3', 071° 22.3'). Regression is for the mean summer chlorophyll *a* values.

¹Graduate School of Oceanography, University of Rhode Island, Narragansett, Rhode Island 02882-1197, USA.

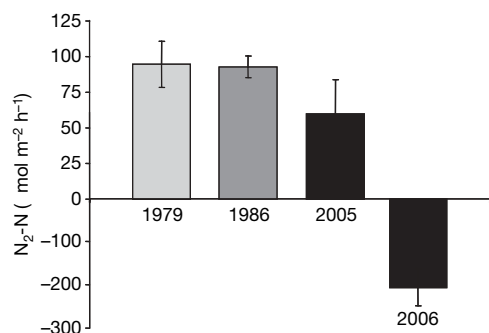


Figure 2 | Mean summer (17–23 °C) N_2 fluxes in mid-bay. Positive rates indicate denitrification, negative rates indicate N-fixation. 1979 data are from ref. 29 (mean \pm s.d., $n = 2$), 1985/86 data are from ref. 30 (mean \pm s.d., $n = 7$), 2005 data are from ref. 7 (mean \pm s.d., $n = 6$), and 2006 data are from this study (mean \pm s.d., $n = 6$). Historical denitrification rates are significantly (one-way ANOVA; $P = 0.002$) higher than those measured in 2005/06 (see the Methods).

lower values for summer blooms²⁴. We assumed a deposition of 25% of four equal blooms based on a June through September carbon fixation of 200 g C m^{-2} measured in mid-bay water during the early 1970s²⁵. Each mesocosm was randomly assigned to one of three levels of organic addition in the form of commercial spray-dried marine phytoplankton with a C:N ratio of 10.5. Beginning in June, three mesocosms received no organic enrichment (0X), three received the spray-dried plankton in four monthly aliquots of 6.25 g C m^{-2} (1/2X) each, and three received 12.5 g C m^{-2} (1X) each. Our intention was that the 50 g C m^{-2} (1X) total summer deposition might approximate the field condition as it was during the 1970s.

We measured the net sediment N_2 flux at the end of the experiment in September 2006. Organic matter loading had a progressive effect on net sediment N_2 fluxes, with denitrification proceeding in the most heavily enriched treatments, both denitrification and nitrogen fixation occurring at relatively low rates in the medium-enrichment treatments, and only nitrogen fixation taking place in the non-enrichment treatments (Fig. 3). Field-collected sediment cores were confirmed to exhibit net nitrogen fixation in early September 2006 at rates similar to the non-enriched mesocosms. After this determination, we added spray-dried plankton to the cores to replicate the 1X enrichment. Two weeks later these cores had also switched from N-fixation ($-250 \mu\text{mol N}_2\text{-N m}^{-2} \text{ h}^{-1}$) to denitrification ($530 \mu\text{mol N}_2\text{-N m}^{-2} \text{ h}^{-1}$). Thus, only a brief exposure to organic matter was needed to reverse the net sediment N_2 flux.

N-fixation associated with cyanobacteria mats and seagrass beds can be an important nitrogen source in shallow, oligotrophic, tropical systems⁵. In contrast, benthic N-fixation has historically accounted for a very small percentage of the total system nitrogen in deeper temperate estuaries and bays (that is 0.1% in Narragansett Bay²⁶ to 4% in Rhode River, Chesapeake Bay⁵). Preliminary analysis of the sediments at the mid-bay station found no cyanobacteria pigments, suggesting that heterotrophic bacteria are responsible for the N-fixation. This is consistent with the fact that all N-fixation incubations were made in complete darkness and the mesocosms were maintained for four months in the dark before the N-fixation measurements.

Table 1 | N-fixation during the summer of 2006 in Narragansett Bay

Site	Latitude Longitude	Temperature (°C)	$N_2\text{-N}$ (\pm s.d.) ($\mu\text{mol m}^{-2} \text{ h}^{-1}$)
1	41° 46.7'	18	-25 (± 3.5)
	071° 22.8'	22	-245 (± 9.7)
2	41° 43.1'	18	-30 (± 5.2)
	071° 21.6'	22	Not measured
3	41° 43.2'	18	-135 (± 19.4)
	071° 18.5'	22	-650 (± 200.4)
4*	41° 35.3'	18	-75 (± 13.2)
	071° 22.3'	22	-276 (± 90.4)

* Historic station discussed in the text.

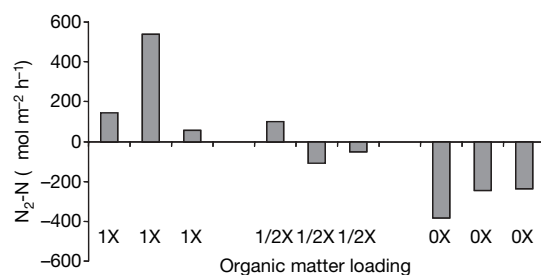


Figure 3 | Net N_2 fluxes measured across a gradient of organic matter loading in mesocosms at the Graduate School of Oceanography. Sediments were collected from the mid-bay historic site. Sediments labelled 1X received a total of 50 g C m^{-2} , 1/2X systems received 25 g C m^{-2} , and 0X systems received no organic matter. Positive rates indicate net denitrification, negative rates indicate net N-fixation.

Simultaneous nitrogen fixation and denitrification have been observed in shallow, subtropical, estuarine sediments²⁷. The N_2/Ar technique measures net N_2 production or uptake, so it is impossible for us to know whether this was the case for these temperate sediments. However, it is possible to determine whether the net N_2 flux is positive (that is removing nitrogen through denitrification) or negative (that is adding nitrogen through N-fixation) to the bay over the annual cycle. If we apply the N-fixation rates observed at the mid-bay site in the summer of 2006 to the entire bay, we estimate that N-fixation introduced $(1.4\text{--}2.2) \times 10^6$ moles of nitrogen per day. This is roughly 1.5 times greater than nitrogen inputs over the summer from the land and atmosphere combined¹⁷. This is a conservative estimate as N-fixation rates measured at the other three stations were higher (Table 1). Because the N-fixation is seasonal, the major nitrogen addition occurred from June through August and amounted to $(125\text{--}200) \times 10^6$ moles of nitrogen during this period. If we apply the mean annual denitrification rate observed during 2005 ($40 \mu\text{mol N}_2\text{-N m}^{-2} \text{ h}^{-1}$) to the whole bay over the remaining nine months of the year, denitrification could remove about 86×10^6 moles of N. Thus, over the complete annual cycle in 2006 the sediments were adding a net $(40\text{--}110) \times 10^6$ moles of nitrogen to the bay. This is equal to 20–60% of the annual nitrogen input from direct sewage discharge.

During summer, dissolved inorganic nitrogen concentrations throughout Narragansett Bay proper are very low and strongly limit primary production²⁸ (Supplementary Fig. 1). In the past, this summer nitrogen limitation was eased by the remineralization of nitrogen deposited on the bottom during the winter–spring bloom. Presumably, the dissolved inorganic nitrogen that used to be taken up by the winter–spring bloom is now flushed offshore. But the climate-induced oligotrophication of the bay has also now made the sediments a new source of reactive nitrogen. It is too soon to know whether this source will become large enough to replace the dissolved inorganic nitrogen being lost offshore during winter. The conversion of the sediments from a significant sink for N_2 over the annual cycle to a large nitrogen source illustrates the complex interplay of climate and biology as they influence biogeochemical cycling in the coastal environment. With these observed changes, a new model for nitrogen dynamics in coastal systems is emerging. Some estuaries may no longer provide the nitrogen retention and removal services they once did. Instead, nitrogen may be exported from the estuary to the continental shelf and, perhaps, to the open ocean, thus shifting the effect of anthropogenic nitrogen loading beyond the immediate coastal zone.

METHODS

Field sampling. Three replicate sediment cores (of 10 cm inner diameter and 30.5 cm long) were collected from the mid-bay site using SCUBA divers. Sediment cores (of 10 cm inner diameter, 30.5 cm long) from the three other sites were collected using a 5 m pull corer onboard the R/V *Zostera*. The cores

were stored at *in situ* temperatures and taken back to the laboratory at the Graduate School of Oceanography of the University of Rhode Island, where they were placed in a water bath and maintained at field temperatures in a dark environmental chamber. The cores were left uncapped with air bubbling gently through the overlying water for about 8–12 h. Before the net N_2 flux incubation the water overlying each core was carefully replaced with filtered (1 μ m) lower Narragansett Bay (~ 32 practical salinity units) water. The cores were then sealed with a gas-tight lid (with no air headspace) and replicate water samples for N_2/Ar analysis were collected at five points over the course of an incubation and preserved with $HgCl_2$. Incubations at 17–23 °C lasted between 6 and 24 h.

Mesocosm experiment. We used a large box corer to collect sediments in May 2006 to a depth of about 30 cm from near the mid-bay station where the long-term chlorophyll measurements are made and close to the historical site of mid-bay benthic fluxes (41° 35.3', 071° 22.3'). We took great care not to disturb the sediments and to preserve their vertical structure. The sediments were maintained in the dark in nine 4 m² mesocosms with 65 cm of overlying water, which was slowly mixed and exchanged at 10% each day with water from the lower West Passage of the bay. The sediments in the field are at 7 m (mean low water, mlw) with a 1.1 m tidal range, receive less than 1–2% of surface illumination, and are completely heterotrophic. In September sediment cores (of 10 cm inner diameter, 30.5 cm long) were collected from each of the mesocosms with a pull corer and transported to the environmental chamber for the N_2/Ar incubation as described above.

Laboratory analysis. Dissolved gas samples were analysed with a quadrupole membrane inlet mass spectrometer (MIMS, Bay Instruments) that only requires a small sample size (<10 ml) and no sample preparation. The instrument provides rapid throughput (~ 20 –30 samples per hour) and a precision of $\pm 0.03\%$ (ref. 22). The N_2/Ar ratio is actually a measure of net N_2 production or consumption (gross denitrification minus gross nitrogen fixation). For the N_2/Ar method, the change in N_2 concentration was determined from the change in the measured N_2/Ar multiplied by the Ar concentration at air saturation. N_2 production for each of the triplicate cores was then determined from a five-point linear regression (Supplementary Fig. 2). Rates were then corrected for the volume of water overlying the core and area of the core. With assistance from M. Hayn at Cornell University and L. Cole from the University of Virginia, we confirmed that N-fixation was occurring at three of the sites, using the more traditional acetylene technique.

Statistical analysis. Differences between the historical and most recently collected sediment oxygen demand, sediment net N_2 fluxes, and the fluxes of ammonium and dissolved inorganic phosphate across the sediment water interface and their relationship with temperature were examined with the statistical analysis program SAS, using a two-way analysis of variance (ANOVA) and least-significant-difference multiple comparison test. All of the fluxes in this study were significantly ($P < 0.01$) lower than those measured previously. Using ANOVA we also tested to see if the historical summer (17–23 °C) N_2 fluxes in mid-bay were different from each other. The historical summer rates are significantly ($P = 0.002$) different from those measured in 2005 and in 2006.

Cyanobacteria analysis. Six surficial (0–2 cm) sediment samples were sent to the laboratory of H. Paerl at the University of North Carolina for cyanobacteria extraction³¹. No cyanobacteria indicator pigments could be found.

Received 26 December 2006; accepted 25 May 2007.

- Howarth, R. W. & Marino, R. Nitrogen as the limiting nutrient for eutrophication in coastal marine ecosystems: Evolving views for over three decades. *Limnol. Oceanogr.* **51**, 364–376 (2006).
- Galloway, J. N. & Cowling, E. B. Nitrogen and the world: 200 years of change. *Ambio* **31**, 64–71 (2002).
- Nixon, S. W. *et al.* The fate of nitrogen and phosphorus at the land-sea margin of the North Atlantic Ocean. *Biogeochemistry* **35**, 141–180 (1996).
- Seitzinger, S. P. & Giblin, A. E. Estimating denitrification in north Atlantic continental shelf sediments. *Biogeochemistry* **35**, 235–260 (1996).
- Howarth, R. W., Marino, R., Lane, J. & Cole, J. J. Nitrogen fixation in freshwater, estuarine, and marine ecosystems. 1. Rates and importance. *Limnol. Oceanogr.* **33**, 669–687 (1988a).
- Li, Y. & Smayda, T. Temporal variability of chlorophyll in Narragansett Bay 1973–1990. *ICES J. Mar. Sci.* **55**, 661–667 (1998).
- Fulweiler, R. W. & Nixon, S. W. Responses of benthic-pelagic coupling to climate change in a temperate estuary. *Hydrobiologia*. (in the press).
- Walther, G. R. *et al.* Ecological response to recent climate change. *Nature* **416**, 389–395 (2002).
- Farnsworth, E. J., Núñez-Farfán, S. A., Careaga, S. A. & Bazzaz, F. A. Phenology and growth of three temperate forest life forms in response to artificial soil warming. *J. Ecol.* **83**, 967–977 (1995).
- Townsend, D. W. & Cammen, L. M. Potential importance of the timing of spring plankton blooms to benthic-pelagic coupling and recruitment of juvenile demersal fishes. *Biol. Oceanogr.* **5**, 215–229 (1988).
- Edwards, M. & Richardson, A. J. Impact of climate change in marine pelagic phenology and trophic mismatch. *Nature* **430**, 881–884 (2004).

- Kemp, W. M. *et al.* Eutrophication of Chesapeake Bay: historical trends and ecological interactions. *Mar. Ecol. Prog. Ser.* **303**, 1–29 (2005).
- Nixon, S. W., Kelly, J. R., Furnas, B. N., Oviatt, C. A. & Hale, S. S. in *Marine Benthic Dynamics* (eds Tenore, K. R. & Coull, B. S.) 219–242 (Univ. of South Carolina Press, Columbia, 1980).
- Nixon, S. W., Granger, S., Buckley, B. A., Lamont, M. & Rowell, B. A one hundred and seventeen year coastal water temperature record from Woods Hole, Massachusetts. *Estuaries* **27**, 397–404 (2004).
- Oviatt, C. A., Keller, A. & Reed, L. Annual primary production in Narragansett Bay with no bay-wide winter-spring phytoplankton bloom. *Estuar. Coast. Shelf Sci.* **54**, 1013–1026 (2002).
- Borkman, G. D. *Analysis and Simulation of Skeletonema costatum* (grev.) Cleve Annual Abundance Patterns in Lower Narragansett Bay 1959–1996 PhD dissertation, Univ. Rhode Island (2002).
- Nixon, S. W. *et al.* in *Ecosystem-Based Estuary Management: A Case Study of Narragansett Bay* (eds Desbonnet, A. Colt, A. B. & Costa-Pierce, B. A.) (Springer, in the press).
- Borkman, G. D. & Smayda, T. J. Long-term trends in water clarity revealed by secchi-disk measurements in lower Narragansett Bay. *ICES J. Mar. Sci.* **55**, 668–679 (1998).
- Brush, M. J., Brawley, J. W., Nixon, S. W. & Kremer, J. N. Modeling phytoplankton production: problems with the Eppley curve and an empirical alternative. *Mar. Ecol. Prog. Ser.* **238**, 31–45 (2002).
- Nixon, S. W., Granger, S. L. & Nowicki, B. L. An assessment of the annual mass balance of carbon, nitrogen, and phosphorus in Narragansett Bay. *Biogeochemistry* **31**, 15–61 (1995).
- Sisler, F. D., & ZoBell, C. E. Nitrogen fixation by sulfate reducing bacteria indicated by nitrogen/argon ratios. *Science* **113**, 511–512 (1951).
- Kana, T. M. *et al.* Membrane inlet mass-spectrometer for rapid high-precision determination of N_2 , O_2 , and Ar in environmental water samples. *Anal. Chem.* **66**, 4166–4170 (1994).
- Galloway, J. N. *et al.* Nitrogen cycles: past, present, and future. *Biogeochemistry* **70**, 153–226 (2004).
- Keller, A. A., Oviatt, C. A., Walker, H. A. & Hawk, J. D. Predicted impacts of elevated temperature on the magnitude of the winter-spring phytoplankton bloom in temperate coastal waters: A mesocosm study. *Limnol. Oceanogr.* **44**, 344–356 (1999).
- Furnas, M. J., Hitchcock, G. L. & Smayda, T. J. in *Estuarine Processes Vol. 1 Uses, Stresses and Adaptation to the Estuary* (ed. Wiley, M.) 118–133 (Academic Press, New York, 1976).
- Seitzinger, S. P. & Garber, J. H. Nitrogen fixation and $^{15}N_2$ calibration of the acetylene reduction assay in coastal marine sediments. *Mar. Ecol. Prog. Ser.* **37**, 65–76 (1987).
- Gardner, W. S. *et al.* Nitrogen fixation and dissimilatory nitrate reduction to ammonium (DNRA) support nitrogen dynamics in Texas estuaries. *Limnol. Oceanogr.* **51**, 558–568 (2006).
- Oviatt, C. A. *et al.* An ecosystem level experiment on nutrient limitation in temperate coastal marine environments. *Mar. Ecol. Prog. Ser.* **116**, 171–179 (1995).
- Seitzinger, S. P. *The Importance of Denitrification and Nitrous Oxide Production in the Nitrogen Dynamics and Ecology of Narragansett Bay*, Rhode Island PhD dissertation, Univ. Rhode Island (1982).
- Nowicki, B. L. *The Fate of Nutrient Inputs to Estuaries: Evidence from Estuarine Mesocosms* PhD dissertation, Univ. Rhode Island (1991).
- Pinckney, J. L. *et al.* Application of photopigment biomarkers for quantifying microalgal community composition and *in situ* growth rates *Org. Geochem.* **32**, 585–595 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This material is based on work supported by the National Science Foundation to the Rhode Island EPSCoR programme, as well as the Rhode Island Sea Grant College Program (NOAA), the Rhode Island Coastal Resource Management Council, and the Switzer Foundation. We acknowledge and appreciate the assistance of A. Giblin, J. Tucker and K. Koop-Jacobsen of the Marine Biological Laboratory and J. Cornwell, M. Owens, T. Kana, and J. Seabrease of the University of Maryland. We also thank H. Paerl and K. Rossignol of the University of North Carolina. In addition we thank M. Hayn at Cornell University and L. Cole at the University of Virginia. This manuscript was improved with comments from A. Giblin, M. E. Q. Pilson and B. Jenkins. For field assistance we thank P. DiMilla, K. Hanks, K. Henry, M. Horn, J. Krumholz, C. Mueller, A. Oczkowski, A. Pimenta, M. Traber, and N. Hovey. We thank L. Harris for assistance with the primary production model.

Author Contributions R.W.F. was responsible for sediment collection as well as sample and data analysis for net N_2 measurements. R.W.F. and S.W.N. co-wrote this manuscript. B.A.B. was responsible for statistical analysis. B.A.B. and S.L.G. supervised the large-scale mesocosm experiment.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to R.W.F. (wally@gso.uri.edu).

Spreading rate dependence of gravity anomalies along oceanic transform faults

Patricia M. Gregg¹, Jian Lin², Mark D. Behn² & Laurent G. J. Montési²

Mid-ocean ridge morphology and crustal accretion are known to depend on the spreading rate of the ridge. Slow-spreading mid-ocean-ridge segments exhibit significant crustal thinning towards transform and non-transform offsets^{1–12}, which is thought to arise from a three-dimensional process of buoyant mantle upwelling and melt migration focused beneath the centres of ridge segments^{1,2,4–7,9,10,12}. In contrast, fast-spreading mid-ocean ridges are characterized by smaller, segment-scale variations in crustal thickness, which reflect more uniform mantle upwelling beneath the ridge axis^{13–15}. Here we present a systematic study of the residual mantle Bouguer gravity anomaly of 19 oceanic transform faults that reveals a strong correlation between gravity signature and spreading rate. Previous studies have shown that slow-slipping transform faults are marked by more positive gravity anomalies than their adjacent ridge segments^{1,2,4,6}, but our analysis reveals that intermediate and fast-slipping transform faults exhibit more negative gravity anomalies than their adjacent ridge segments. This finding indicates that there is a mass deficit at intermediate- and fast-slipping transform faults, which could reflect increased rock porosity, serpentinization of mantle peridotite, and/or crustal thickening. The most negative anomalies correspond to topographic highs flanking the transform faults, rather than to transform troughs (where deformation is probably focused and porosity and alteration are expected to be greatest), indicating that crustal thickening could be an important contributor to the negative gravity anomalies observed. This finding in turn suggests that three-dimensional magma accretion may occur near intermediate- and fast-slipping transform faults.

We analyse the residual mantle Bouguer gravity anomaly (RMBA) from 19 oceanic transform fault systems and their adjacent ridge segments at ultraslow- to fast-spreading mid-ocean ridges (Supplementary Fig. 1) with fault slip rates of 1.3–14.5 cm yr^{−1} and fault lengths of 70–350 km (see Methods). RMBA is used to infer local density variations relative to a reference model that includes a 6-km-thick crust. A negative RMBA indicates a mass deficit, which can be due to thickened crust, increased porosity, and/or serpentinized mantle. Conversely, a positive RMBA typically implies crustal thinning. Our results reveal a systematic disparity in the gravity signature between ultraslow-/slow-slipping transform faults and intermediate-/fast-slipping transform faults.

To illustrate these differences, we compare the RMBA at the fast-slipping Siqueiros transform to the RMBA at the slow-slipping Atlantis transform (Fig. 1). The 150-km-long Siqueiros transform located at ~8.3° N on the East Pacific Rise is slipping at ~11.8 cm yr^{−1} and includes five fault segments connected by four intra-transform spreading centres¹⁶ (Fig. 1a). The ridge segments to the north and south of Siqueiros are typical of fast-spreading ridges with axial highs as shallow as 2.0–2.8 km. The ridge transform

intersections bounding Siqueiros display overshooting ridge tips¹⁶ typical of ridge transform intersections at magmatically robust ridge segments (Supplementary Figs 2–5). In contrast, the 70-km-long Atlantis transform at 30° N on the Mid-Atlantic Ridge is slipping at ~2.4 cm yr^{−1} and contains a single strike-slip fault zone (Fig. 1b). The ridge segments adjacent to the Atlantis transform display slow-spreading rift valleys and a wide range of axial depths.

As seen in previous gravity studies of the Atlantis transform^{2,7,17}, more negative RMBA is associated with adjacent ridge segment centres (up to ~80 mGal less than the transform offset), whereas the most positive RMBA is located in the transform fault and fracture zone (Fig. 1d). In contrast, at Siqueiros more positive RMBA is observed along the adjacent ridge segments and more negative values are found within the transform fault and fracture zone domains (up to ~40 mGal less than the adjacent ridge segments). RMBA minima often correspond to bathymetric highs along the eastern and southern flanks of the fault (Fig. 1c).

The difference in the average RMBA values between transform fault domains and their adjacent ridge segments— $\Delta\text{RMBA}_{\text{T}-\text{R}}$ —varies systematically with spreading rate for the 19 transform faults included in this study (Fig. 2, see Methods section for $\Delta\text{RMBA}_{\text{T}-\text{R}}$ calculation). $\Delta\text{RMBA}_{\text{T}-\text{R}} > 0$ implies more positive RMBA along the transform fault domain than along the adjacent ridge segments, whereas $\Delta\text{RMBA}_{\text{T}-\text{R}} < 0$ implies more negative values within the transform fault domain. Along ultraslow- and slow-spreading ridges, the average RMBA in transform fault domains is systematically more positive than along the adjacent ridge segments with $\Delta\text{RMBA}_{\text{T}-\text{R}}$ of up to 45 mGal. However, for spreading rates exceeding ~5 cm yr^{−1} the average RMBA within transform fault domains becomes more negative than on the adjacent ridge segments. Most intermediate- and fast-slipping transforms are associated with $\Delta\text{RMBA}_{\text{T}-\text{R}}$ of −4 to −14 mGal, whereas the Blanco transform is associated with a $\Delta\text{RMBA}_{\text{T}-\text{R}}$ of −40 mGal (Fig. 2 and Supplementary Table 1).

Previous studies attribute positive RMBA along slow-slipping transform faults to decreased crustal accretion towards the end of ridge segments and therefore crustal thinning within the transform fault domain^{1,2,4,6}. If the negative RMBA observed at intermediate- and fast-slipping transforms also reflects variations in crustal thickness, these anomalies indicate crustal thickening within the transform domain. However, negative RMBA can also arise from other factors, such as cracking and alteration at transform fault offsets^{5,8–10}. Therefore, we use two-dimensional (2D) gravity models to estimate the amplitude of $\Delta\text{RMBA}_{\text{T}-\text{R}}$ corresponding to: (1) increased rock porosity; (2) serpentinization of mantle peridotite; and (3) crustal thickening (Fig. 3, Supplementary Fig. 7, and Supplementary Methods).

Low seismic velocities recorded in the Clipperton transform and near the Siqueiros transform have been attributed to increased cracking and porosity within the transform domain^{15,18}. The observed

¹Department of Geology and Geophysics, Massachusetts Institute of Technology/Woods Hole Oceanographic Institution Joint Program, ²Department of Geology and Geophysics, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA.

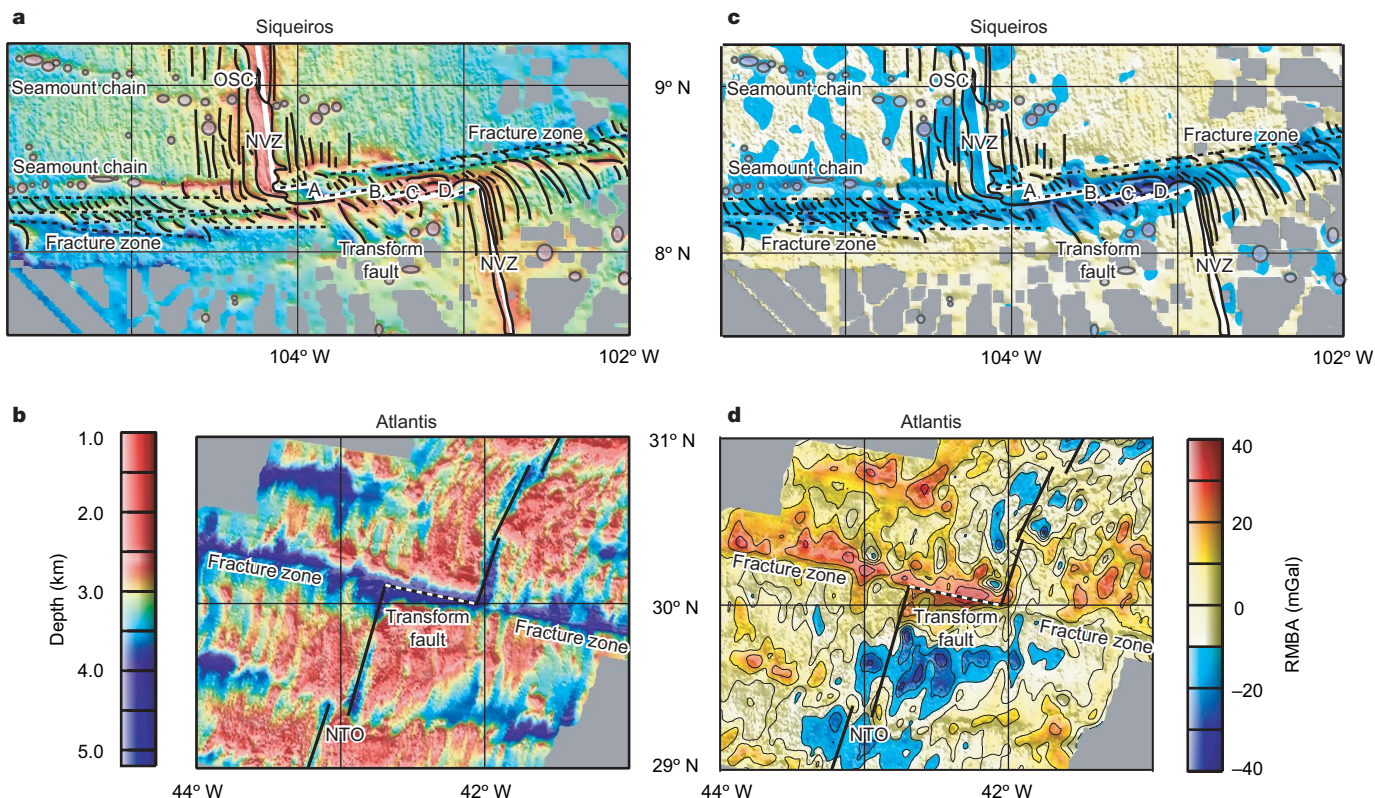


Figure 1 | Comparison of bathymetry and RMBA of the Siqueiros transform on the East Pacific Rise and Atlantis transform on the Mid-Atlantic Ridge, at the same map scale. a, Bathymetric map for the Siqueiros transform with geologic interpretation¹⁶. Solid black lines indicate the seafloor fabric, circles show locations of seamounts, and dashed black lines show the locations of the fracture zones and transform fault. The solid white line indicates the location of the plate boundary used in the 3D thermal model. ITSCs are labelled A, B, C and D¹⁶. NVZ, neo-volcanic zone; OSC, overlapping

spreading centre. **b**, Bathymetric map for the Atlantis transform. The solid black line indicates location of the plate boundary used in the 3D thermal model, and the dashed portion indicates the Atlantis transform fault. **c**, Calculated RMBA map for the Siqueiros transform fault with geologic interpretation overlaid from **a**. **d**, Calculated RMBA map for the Atlantis transform fault. Seabeam bathymetry compiled by the RIDGE multibeam synthesis project (<http://ocean-ridge.ldeo.columbia.edu/general/html/home.html>).

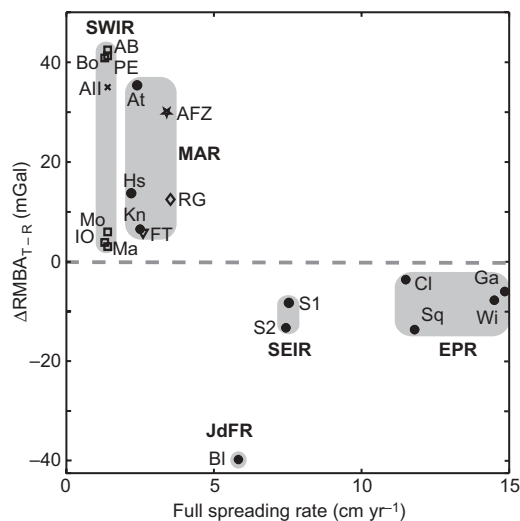


Figure 2 | Compilation of ΔRMBA_{T-R} values for the 19 transform systems analysed. The new RMBA calculations from this study are marked with solid circles. Previous RMBA studies are indicated by the open symbols: square²⁹, triangle¹², cross³⁰, star¹¹, diamond³. Abbreviations for ridges: SWIR, Southwest Indian Ridge; MAR, Mid-Atlantic Ridge; JdFR, Juan de Fuca Ridge; SEIR, Southeast Indian Ridge; EPR, East Pacific Rise. Abbreviations for transform faults: Bo, Bouvet; IO, Islas Orcadas; AB, Andrew Bain; PE, Prince Edward; All, Atlantis II; Mo, Moshesh; Ma, Marion; At, Atlantis; Hs, Hayes; Kn, Kane; FT, Fifteen-Twenty; AFZ, Ascension; RG, Rio Grande; BI, Blanco; S2, SEIR2; S1, SEIR1; Cl, Clipperton; Sq, Siqueiros; Wi, Wilkes; Ga, Garrett.

decrease in seismic velocities of $\sim 1 \text{ km s}^{-1}$ in the Clipperton transform fault¹⁸ can be explained by porosity increases in the range of 1–5.5% for crack aspect ratios of 0.01–0.1 (ref. 19). Calculations in Fig. 3 demonstrate that an average porosity increase of 2–7% within the upper crust down to the brittle-ductile transition (600°C) is required to produce the negative ΔRMBA_{T-R} values of up to -14 mGal observed at fast-slipping transforms (Fig. 2). However,

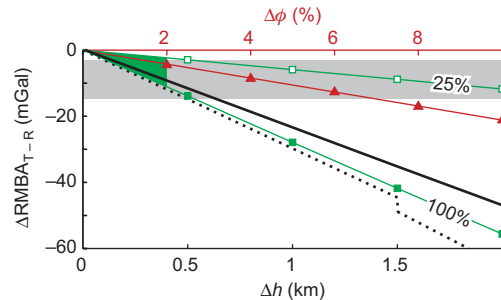


Figure 3 | Results of 2D forward models showing the predicted ΔRMBA_{T-R} . Data are shown as a function of increase in crustal thickness through layer 2A (dotted black line) or overall crustal thickening (solid black line); increases in porosity (ϕ , red line); or the presence of serpentinized mantle (green lines indicate percentage; green shading indicates serpentine stability field). Δh indicates change in layer thickness. Grey shading indicates the observed ΔRMBA_{T-R} at fast-slipping transforms. The kink in the curve for increase in layer 2A corresponds to thickening of layer 2A beyond the thickness of layer 2B in the reference model. See Supplementary Fig. 7 for model details.

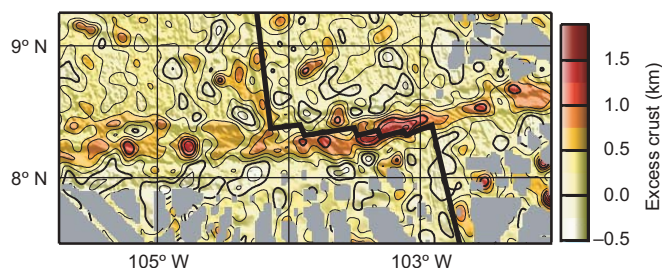


Figure 4 | Lateral variations in crustal thickness required to explain the observed RMBA in the Siqueiros transform system (Fig. 1c). Excess crust is defined as the deviation from the reference crustal thickness of 6 km, assuming sea water, crust and mantle densities of 1,030, 2,730 and 3,300 kg m⁻³, respectively.

an increase in porosity in excess of 5% at slow-slipping transforms will negate the effect of crustal thinning and result in negative $\Delta\text{RMBA}_{\text{T-R}}$ (Supplementary Fig. 7). Consequently, a porosity increase of 2–5% is consistent with observed $\Delta\text{RMBA}_{\text{T-R}}$ and seismic data at both slow- and fast-slipping transform faults.

Serpentinization of mantle peridotite can also generate negative $\Delta\text{RMBA}_{\text{T-R}}$. However, the calculated 500 °C isotherm, which represents the upper limit of serpentine stability²⁰, is located within the crust at all fast-slipping transform faults considered in this study (see Methods) and therefore predicts no mantle serpentinization. Invoking hydrothermal cooling (see Supplementary Methods) depresses the 500 °C isotherm to an average depth of 6.4 km at Siqueiros transform fault. Assuming 100% serpentinization (density of 2,550 kg m⁻³) of the mantle where temperatures are <500 °C results in a negative $\Delta\text{RMBA}_{\text{T-R}}$ of 0 to –10 mGal, which is slightly less than the values of up to –14 mGal observed at fast-slipping transform faults (Fig. 3). However, if 100% serpentinization were pervasive above the 500 °C isotherm at slow-slipping transform faults, it would negate the effects of crustal thinning and produce negative $\Delta\text{RMBA}_{\text{T-R}} < -100$ mGal (Supplementary Fig. 7g). Therefore, although low degrees of mantle serpentinization might contribute to the observed negative RMBA along fast-slipping transform faults, current thermal models for oceanic transform faults make it unlikely that serpentinization alone can explain the systematic differences between fast- and slow-slipping transforms.

Finally, we estimate the magnitude of variations in crustal thickness required to explain the RMBA observed along the Siqueiros transform by inverting the observed RMBA relative to a reference model with average crustal density of 2,730 kg m⁻³ (ref. 1) (Fig. 4). Along the eastern domain of Siqueiros transform, we predict an excess crustal thickness of >1.5 km, while in the western domain of the transform and in the off-axis fracture zones excess crust ranges from 0.5–1.5 km. The $\Delta\text{RMBA}_{\text{T-R}}$ observed at fast-slipping transform faults can also be produced by increasing layer 2A by 0.25–0.6 km (Fig. 3).

Although increased rock porosity and mantle serpentinization could both contribute to the observed negative RMBA along intermediate- and fast-slipping transform faults, we believe that local crustal thickening is probably the dominant mechanism. Areas of localized deformation, such as fault zones, are expected to exhibit enhanced cracking and alteration. However, the most negative RMBA values observed at intermediate- and fast-slipping transform faults correspond to bathymetric highs flanking the transforms instead of the transform troughs where deformation is likely to be localized. Local excess crust within intermediate- and fast-slipping transform fault domains may result from a combination of magmatic accretion mechanisms including: (1) lateral transport of excess magma from the ridge axis into the transform fault domain^{16,21}; (2) magmatic accretion at intra-transform spreading centres^{16,22}; (3) the pooling of lava within topographic lows²³; and/or (4) 'leaky' magmatic accretion along the entire transform fault²⁴.

Lateral dyke propagation from a central magma chamber has been proposed to explain crustal thickness variations along slow-spreading ridge segments²⁵. At magmatically robust segments along intermediate and fast-spreading ridges, a continuous magma chamber may exist beneath the entire segment and feed dykes from the ridge transform intersection into the transform fault domain. A possible geologic expression of this mechanism is overshooting ridge tips, which are common bathymetric features of robust ridge segments at intermediate- and fast-spreading ridge transform intersections¹⁶ (see Supplementary Figs 2–5). Bathymetry and RMBA of overshooting ridge tips at intermediate and fast-spreading ridges suggest that dykes propagating along the ridge axis penetrate past the transform fault into juxtaposed old oceanic crust, curve in the direction of the transform, and locally increase the crustal thickness of an already full-thickness crust (Fig. 5b). Geologic evidence from the northwestern segment of Blanco transform indicates significant

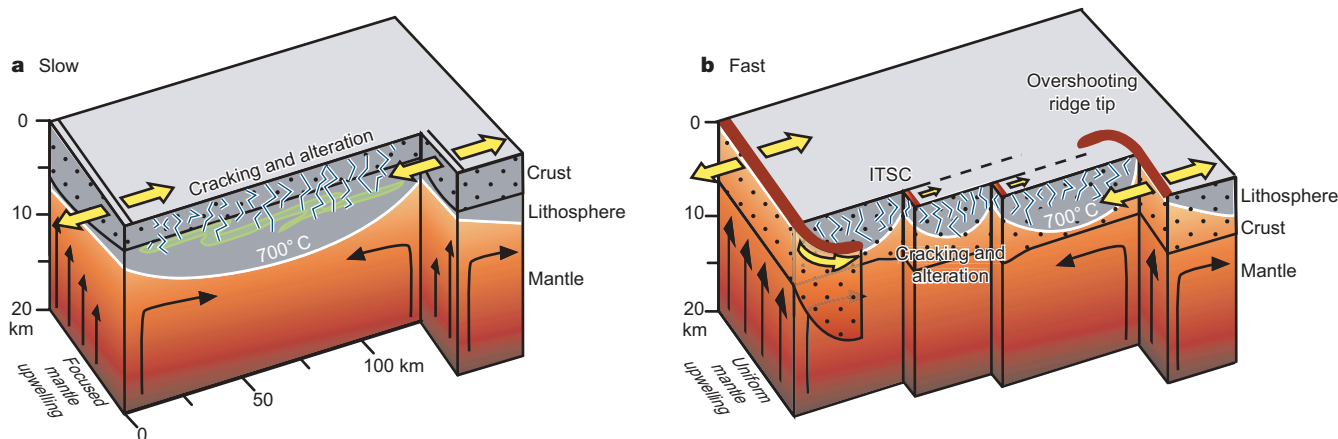


Figure 5 | A spreading-rate-dependent model of crustal accretion and mantle upwelling based on observed RMBA calculations and morphological features at transform fault systems on slow- and fast-spreading ridges. Large yellow arrows indicate plate motion vectors; smaller black arrows indicate mantle flow. The grey area with white outline represents the lithosphere as defined by the 700 °C isotherm, while the crust is identified by dotted pattern. The blue cracks and the green wiggles indicate regions of increased porosity and serpentinization, respectively. **a**, Slow-spreading

ridges exhibit focused mantle upwelling at segment centres and crustal thinning towards segment ends and transform faults. **b**, Fast-spreading ridges have relatively uniform mantle upwelling. The negative $\Delta\text{RMBA}_{\text{T-R}}$ at fast-spreading ridges may result from: (1) excess magma supply on ridge segments causing over-shooting ridges and dyke injection into the transform fault domain; (2) increased porosity within the transform; (3) actively accreting ITSCs; (4) pooling of extruded lavas within the transform fault.

increases in the extrusive lava layer with a possible overshooting ridge tip source²¹. However, for this model to explain the RMBA along the entire transform fault (rather than just in the fracture zone), transform-parallel dyke propagation must extend at least half the distance between ridge transform intersections (~50–75 km). Bathymetric data demonstrate that overshooting ridge tips wrap around and into transform faults, but the cause of this curvature is unclear. Possible explanations include response to the local stress field at the ridge transform intersection, and/or dyke propagation along preferential pathways created by pre-existing transform-parallel cracks.

Fast-slipping transform faults typically contain multiple fault segments separated by intra-transform spreading centres (ITSC)^{13,26}. Several ITSCs at East Pacific Rise transform faults have been sampled and contain fresh basaltic lavas indicating crustal accretion within the transform fault domain^{16,22}. In some cases, such as the Garrett transform fault (Supplementary Fig. 2), the ITSCs occur in pull-apart basins and do not appear to increase the crustal thickness. Alternatively, along the eastern portion of Siqueiros transform (Fig. 1a) ITSCs are associated with negative RMBA. Thus, ITSCs may thicken crust in some localities.

A third possible mechanism for local crustal thickening is the pooling of erupted lavas in topographic lows within the transform fault domain²³ (Fig. 5b). However, the majority of the negative RMBA values we have observed correspond with topographic highs, indicating that this might not be a dominant mechanism. Finally, because off-axis melt is considered abundant at fast-spreading ridges²⁷, 'leaky' magmatic accretion might occur along the entire length of some transform faults²⁴. This mechanism could cause increases in crustal thickness and may be enhanced by trans-tensional forces acting upon a transform fault owing to plate motion reorganizations²⁴.

Thus the observed negative RMBA values at intermediate- and fast-slipping transform faults may be the result of a combination of mechanisms, including increases in rock porosity, mantle serpentinization and/or crustal thickening. Further detailed constraints are required to distinguish among the potential contributors to the observed gravity anomalies. However, the strong correlation of the negative RMBA with topographic highs flanking the transforms suggests crustal thickening as a potentially important mechanism for at least some of the observed negative RMBA values. Systematic occurrence of crustal thickening, if confirmed, would require a modification of our current understanding of how three-dimensional magma accretion occurs along intermediate- and fast-spreading ridges in the vicinity of transform fault discontinuities.

METHODS SUMMARY

The mantle Bouguer anomaly was calculated following the method of ref. 1. This technique removes the effect of seafloor topography and 6 km of reference crust from the satellite free air gravity²⁸. The RMBA was calculated by removing from the mantle Bouguer anomaly the effects of lithospheric cooling as estimated from a three-dimensional (3D) passive mantle upwelling model calculated using the commercial finite-element modelling software package COMSOL MULTIPHYSICS 3.2 (for model details see Supplementary Tables 1 and 2). The RMBA reflects the deviation of sub-seafloor density structure from the reference model described in Supplementary Table 2. When constructing the thermal models for segmented transform faults, we incorporated the best-known geometry of transform fault segments and assumed steady-state crustal accretion at the ITSCs.

We define $\Delta\text{RMBA}_{T-R} = [(\text{RMBA}_T - \text{RMBA}_{R1}) + (\text{RMBA}_T - \text{RMBA}_{R2})]/2$, where RMBA_{R1} and RMBA_{R2} are the average RMBA values on the adjacent ridge segments 1 and 2, respectively, and RMBA_T is the average RMBA value along the transform fault domain (see Supplementary Fig. 8). Where digital data are available^{12,29}, RMBA_{R1} and RMBA_{R2} were calculated by averaging the values at the centre of each adjacent ridge segment within a 20-km-wide and 50-km-long box. At the East Pacific Rise, Juan de Fuca Ridge, and Southeast Indian Ridge, where the transform fault zones are wider owing to segmentation, RMBA_T was calculated by averaging over a 20-km-wide swath centred 20 km from the ridge transform intersections. At the Mid-Atlantic Ridge and Southwest Indian Ridge the RMBA values were averaged over a 10-km-wide swath centred 20 km from the

ridge transform intersections. Where digital data are unavailable^{3,11,30}, average values were estimated from published profiles and maps of the transform faults and their adjacent ridge segments.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 22 September 2006; accepted 24 May 2007.

1. Kuo, B. Y. & Forsyth, D. W. Gravity anomalies of the ridge transform intersection system in the South Atlantic between 31 and 34.5°S: Upwelling centers and variations in crustal thickness. *Mar. Geophys. Res.* **10**, 205–232 (1988).
2. Lin, J., Purdy, G. M., Schouten, H., Sempere, J.-C. & Zervas, C. Evidence from gravity data for focused magmatic accretion along the Mid-Atlantic Ridge. *Nature* **344**, 627–632 (1990).
3. Blackman, D. K. & Forsyth, D. W. Isostatic compensation of tectonic features of the Mid-Atlantic Ridge: 25°–27°30'S. *J. Geophys. Res.* **96**, 11741–11758 (1991).
4. Lin, J. & Phipps Morgan, J. The spreading rate dependence of three-dimensional mid-ocean ridge gravity structure. *Geophys. Res. Lett.* **19**, 13–16 (1992).
5. Tolstoy, M., Harding, A. & Orcutt, J. Crustal thickness on the Mid-Atlantic Ridge: bulls-eye gravity anomalies and focused accretion. *Science* **262**, 726–729 (1993).
6. Detrick, R. S., Needham, H. D. & Renard, V. Gravity anomalies and crustal thickness variations along the Mid-Atlantic Ridge between 33°N and 40°N. *J. Geophys. Res.* **100**, 3767–3787 (1995).
7. Escartin, J. & Lin, J. Ridge offsets, normal faulting, and gravity anomalies of slow spreading ridges. *J. Geophys. Res.* **100**, 6163–6177 (1995).
8. Minshull, T. A. Along-axis variations in oceanic crustal density and their contribution to gravity anomalies at slow-spreading ridges. *Geophys. Res. Lett.* **23**, 849–852 (1996).
9. Canales, J. P., Detrick, R. S., Lin, J., Collins, J. A. & Toomey, D. R. Crustal and upper mantle seismic structure beneath the rift mountains and across a nontransform offset at the Mid-Atlantic Ridge (35°N). *J. Geophys. Res.* **105**, 2699–2719 (2000).
10. Hooft, E. E. E., Detrick, R. S., Toomey, D. R., Collins, J. A. & Lin, J. Crustal thickness and structure along three contrasting spreading segments of the Mid-Atlantic Ridge, 33.5°–35°N. *J. Geophys. Res.* **105**, 8205–8226 (2000).
11. Bruguier, N. J., Minshull, T. A. & Brozena, J. M. Morphology and tectonics of the Mid-Atlantic Ridge, 7°–12°S. *J. Geophys. Res.* **108**, 2093, doi:10.1029/2001JB001172 (2003).
12. Fujiwara, T. et al. Crustal evolution of the Mid-Atlantic Ridge near the Fifteen-Twenty Fracture Zone in the last 5 Ma. *Geochem. Geophys. Geosyst.* **4**, 1024, doi:10.1029/2002GC000364 (2003).
13. Fox, P. J. & Gallo, D. G. A tectonic model for ridge-transform-ridge plate boundaries: Implications for the structure of oceanic lithosphere. *Tectonophysics* **104**, 205–242 (1984).
14. Macdonald, K. C. et al. A new view of the mid-ocean ridge from the behavior of ridge-axis discontinuities. *Nature* **335**, 217–225 (1988).
15. Canales, J. P., Detrick, R., Toomey, D. R. & Wilcock, S. D. Segment-scale variations in the crustal structure of 150–300 kyr old fast spreading oceanic crust (East Pacific Rise, 8°15'N–10°5'N) from wide-angle seismic refraction profiles. *Geophys. J. Int.* **152**, 766–794 (2003).
16. Fornari, D. J. et al. Structure and topography of the Siqueiros transform-fault system—Evidence for the development of intra-transform spreading centers. *Mar. Geophys. Res.* **11**, 263–299 (1989).
17. Blackman, D. K., Cann, J. R., Janssen, B. & Smith, D. K. Origin of extensional core complexes: Evidence from the Mid-Atlantic Ridge at Atlantis Fracture Zone. *J. Geophys. Res.* **103**, 21315–21333 (1998).
18. Van Avendonk, H. J. A., Harding, A. J., Orcutt, J. A. & McClain, J. S. Contrast in crustal structure across the Clipperton transform fault from travel time tomography. *J. Geophys. Res.* **106**, 10961–10981 (2001).
19. Wilkens, R. H., Fryer, G. J. & Karsten, J. Evolution of porosity and seismic structure of upper oceanic-crust—Importance of aspect ratios. *J. Geophys. Res.* **96**, 17981–17995 (1991).
20. Ulmer, P. & Trommsdorff, V. Serpentine stability to mantle depths and subduction-related magmatism. *Science* **268**, 858–861 (1995).
21. Karson, J. A., Tivey, M. A. & Delaney, J. R. Internal structure of uppermost oceanic crust along the Western Blanco Transform Scarp: Implications for subaxial accretion and deformation at the Juan de Fuca Ridge. *J. Geophys. Res.* **107**, 2181, doi:10.1029/2000JB000051 (2002).
22. Perfit, M. R. et al. Recent volcanism in the Siqueiros transform fault: Picritic basalts and implications for MORB magma genesis. *Earth Planet. Sci. Lett.* **141**, 91–108 (1996).
23. Begnaud, M. L., McClain, J. S., Barth, G. A., Orcutt, J. A. & Harding, A. J. Velocity structure from forward modeling of the eastern ridge-transform intersection area of the Clipperton Fracture Zone, East Pacific Rise. *J. Geophys. Res.* **102**, 7803–7820 (1997).
24. Menard, H. W. & Atwater, T. Origin of fracture zone topography. *Nature* **222**, 1037–1040 (1969).
25. Fialko, Y. A. & Rubin, A. M. Thermodynamics of lateral dike propagation: Implications for crustal accretion at slow spreading mid-ocean ridges. *J. Geophys. Res.* **103**, 2501–2514 (1998).

26. Gregg, P. M., Lin, J. & Smith, D. K. Segmentation of transform systems on the East Pacific Rise: Implications for earthquake processes at fast-slipping oceanic transform faults. *Geology* **34**, 289–292 (2006).
27. Sohn, R. A. & Sims, K. W. W. Bending as a mechanism for triggering off-axis volcanism on the East Pacific Rise. *Geology* **33**, 93–96 (2005).
28. Smith, W. H. F. & Sandwell, D. T. Global seafloor topography from satellite altimetry and ship depth soundings. *Science* **277**, 1957–1962 (1997).
29. Georgen, J. E., Lin, J. & Dick, H. J. B. Evidence from gravity anomalies for interactions of the Marion and Bouvet hotspots with the Southwest Indian Ridge: Effects of transform offsets. *Earth Planet. Sci. Lett.* **187**, 283–300 (2001).
30. Muller, M. R., Minshull, T. A. & White, R. S. Crustal structure of the Southwest Indian Ridge at the Atlantis II Fracture Zone. *J. Geophys. Res.* **105**, 25809–25828 (2000).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by a National Science Foundation (NSF) Graduate Research Fellowship (P.M.G.), the NSF (M.D.B.), and the Woods Hole Oceanographic Institution Deep Ocean Exploration Institute (J.L. and L.G.J.M.). We are grateful for discussions with J. P. Canales, A. Cruse, H. J. B. Dick, D. Fornari, D. Forsyth, J. Georgen, J. Gregg, T. Grove, G. Hirth, D. Lizarralde, J. McGuire, M. Perfit, H. Schouten, D. Smith and the WHOI geophysics group. This manuscript was greatly improved by a review by R. Buck.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to P.G. (pgregg@whoi.edu).

METHODS

Because there is an inherent non-uniqueness in gravity calculations, we have created several 2D forward models³¹ to quantify the effects of various mechanisms on the gravity signature of a ridge–transform system (Supplementary Fig. 7). The light blue and orange shaded regions in Supplementary Fig. 7 indicate the extent of $\Delta\text{RMBA}_{\text{T-R}}$ observed along transform faults at slow- and fast-spreading ridges, respectively. The density used for each of the layers is indicated in Supplementary Fig. 7. These values do vary in nature^{32–35}.

In the first suite of models we varied the crustal thickness by Δh either entirely as a change of layer 2A, layer 2B, or proportionally throughout the entire crust (Supplementary Figs 7a, b and c, respectively). All three mechanisms act to decrease $\Delta\text{RMBA}_{\text{T-R}}$ within the range of fast-slipping transforms with <1 km of excess material. Increasing layer 2A requires the least amount of excess material with 0.25–0.6 km. The kink in model A results from increasing layer 2A beyond the thickness of layer 2B.

The second suite of models considers increased porosity in the brittle layer (temperatures <600 °C). Two 3D thermal models were calculated to obtain the average depth to the 600 °C isotherm along the Siqueiros transform fault: (1) the initial 3D passive mantle upwelling model, outlined in the Methods (the red dashed line at ~ 3 km depth on Supplementary Figs 7d and e), and (2) a hydrothermally cooled model in which heat conductivity is increased tenfold to 4 km depth (the blue dashed line at 6.4 km depth on Supplementary Figs 7d and e). Density changes due to variations in porosity are defined as: $\rho(\Phi) = (1 - \Phi)\rho_{\text{rock}} + \Phi\rho_{\text{sw}}$, where Φ is the porosity, ρ_{rock} is the density of the host rock, and ρ_{sw} is the density of sea water. Assuming that the brittle–ductile transition is at ~ 3 km, porosity would need to be increased 2–7% to achieve the $\Delta\text{RMBA}_{\text{T-R}}$ values of magnitude ~ 14 mGal observed at fast-slipping transform faults.

Similar calculations were performed for slow-slipping transforms, assuming a thicker brittle layer, as was calculated for the Atlantis transform fault. These results indicate that it is difficult to explain the $\Delta\text{RMBA}_{\text{T-R}}$ at both fast and slow-slipping transforms simultaneously, assuming similar porosity structures in both regions. Because it is unlikely that there will be large increases in porosity down to 9 km depth at slow-slipping transform faults, even though it is within the pressure constraints suggested by ref. 36, we have rerun the 2D slow-slipping transform model with increased porosity to 6 km depth. While porosity increases in excess of 8% still act to negate the effect of crustal thinning at slow-slipping transform faults, moderate increases of 2–5% are consistent with $\Delta\text{RMBA}_{\text{T-R}}$ values at both slow- and fast-slipping transform faults (Supplementary Fig. 7e).

The final set of forward models includes serpentinization of the mantle beneath the transform fault (Supplementary Figs 7f and g). As in the porosity models, we have incorporated two 3D thermal calculations at the Siqueiros transform to obtain average depths to the 500 °C isotherm, which limits the serpentine stability field²⁰ to 2.5 km for uncooled and 6.4 km for hydrothermally cooled lithosphere. Without hydrothermal cooling the 500 °C isotherm lies well within the upper crust and there is no predicted mantle serpentinization. However, if hydrothermal cooling is applied from 0 to 4 km depth, an end member model of 100% serpentinization will result in a $\Delta\text{RMBA}_{\text{T-R}}$ of ~ 10 mGal, which is the average of the observed $\Delta\text{RMBA}_{\text{T-R}}$ for fast-slipping transform faults. If a similar model for mantle serpentinization is applied at a slow-slipping transform fault, mantle serpentinization down to the hydrothermally cooled 500 °C isotherm would produce such a large negative RMBA that it would completely offset the effect of crustal thinning, which is inconsistent with observations.

Additionally, we calculated a modified RMBA at the Siqueiros transform using the hydrothermally cooled thermal model and the density for 100% serpentine ($2,550 \text{ kg m}^{-3}$) in all regions of the mantle (>6 km depth) where the temperature is <500 °C. Even though this correction maximizes the effect of serpentine, it decreases the average $\Delta\text{RMBA}_{\text{T-R}}$ within the Siqueiros transform fault domain by only <2 mGal, and reduces the calculated excess crustal thickness by <200 m.

Laboratory experiments suggest that, if the crust is thinned and mantle upwelling occurs, hydrothermal fluid penetration through thermal cracking and grain boundary wetting will reach only ~ 4 – 6 km depth³⁷. Therefore, it is unlikely that hydrothermal fluids will penetrate to great enough depths to serpentinize large regions of mantle beneath fast-slipping transform faults. Moreover, depressing the 500 °C isotherm through hydrothermal cooling is inconsistent with calculated earthquake depths at oceanic transform faults³⁸. On the other hand, slow-slipping transform faults display significant crustal thinning and are cooler than intermediate- and fast-slipping transform faults, implying that a larger portion of the underlying mantle resides in the serpentine stability field. However, the locally positive RMBA observed at slow-slipping transform faults reflects crustal thinning despite the potential for significant serpentinized peridotite⁹. Together, these results indicate that, while mantle serpentinization may

occur at intermediate and fast-slipping transform faults, the direct contribution of serpentinization to the gravity signature is likely to be negligible.

31. Blakely, R. J. *Potential Theory in Gravity and Magnetic Applications* 378 (Cambridge Univ. Press, New York, 1996).
32. Cochran, J. R., Fornari, D. J., Coakley, B. J., Herr, R. & Tivey, M. A. Continuous near-bottom gravity measurements made with a BGM-3 gravimeter in DSV Alvin on the East Pacific Rise crest near 9° 31' N and 9° 50' N. *J. Geophys. Res.* **104**, 10841–10861 (1999).
33. Stevenson, J. M., Hildebrand, J. A., Zumberge, M. A. & Fox, C. G. An ocean-bottom gravity study of the southern Juan-De-Fuca Ridge. *J. Geophys. Res.* **99**, 4875–4888 (1994).
34. Johnson, H. P., Pruis, M. J., Van Patten, D. & Tivey, M. A. Density and porosity of the upper oceanic crust from seafloor gravity measurements. *Geophys. Res. Lett.* **27**, 1053–1056 (2000).
35. Escartin, J., Hirth, G. & Evans, B. Nondilatant brittle deformation of serpentinites: Implications for Mohr–Coulomb theory and the strength of faults. *J. Geophys. Res.* **102**, 2897–2914 (1997).
36. Behn, M. D. & Kelemen, P. B. Relationship between seismic P-wave velocity and the composition of anhydrous igneous and meta-igneous rocks. *Geochim. Geophys. Geosyst.* **4**, 1041, doi:10.1029/2002GC000393 (2003).
37. deMartin, B., Hirth, G. & Evans, B. in *Mid-Ocean Ridges: Hydrothermal Interactions Between Lithosphere and Oceans* (eds German, C. R., Lin, J. & Parsons, L. M.) 318 (American Geophysical Union, Washington DC, 2004).
38. Abercrombie, R. E. & Ekstrom, G. Earthquake slip on oceanic transform faults. *Nature* **410**, 74–77 (2001).

LETTERS

Biodiversity and ecosystem multifunctionality

Andy Hector¹ & Robert Bagchi^{1†}

Biodiversity loss can affect ecosystem functions and services^{1–4}. Individual ecosystem functions generally show a positive asymptotic relationship with increasing biodiversity, suggesting that some species are redundant^{5–8}. However, ecosystems are managed and conserved for multiple functions, which may require greater biodiversity. Here we present an analysis of published data from grassland biodiversity experiments^{9–11}, and show that ecosystem multifunctionality does require greater numbers of species. We analysed each ecosystem function alone to identify species with desirable effects. We then calculated the number of species with positive effects for all possible combinations of functions. Our results show appreciable differences in the sets of species influencing different ecosystem functions, with average proportional overlap of about 0.2 to 0.5. Consequently, as more ecosystem processes were included in our analysis, more species were found to affect overall functioning. Specifically, for all of the analysed experiments, there was a positive saturating relationship between the number of ecosystem processes considered and the number of species influencing overall functioning. We conclude that because different species often influence different functions, studies focusing on individual processes in isolation will underestimate levels of biodiversity required to maintain multifunctional ecosystems.

The potential for diversity to affect ecosystem processes (or functions) was recognized by mid-twentieth-century researchers¹², and originally dates back to Darwin^{13,14}. The recent interest in this topic is due to the realization that the sustained loss of biodiversity could affect ecosystem functioning and the services derived by humans^{1–4}. Meta-analysis of the results of the first generation⁵ of experimental research on biodiversity and ecosystem functioning has revealed that individual ecosystem processes generally show a positive but saturating relationship with increasing diversity, although the mechanisms underlying these relationships are still under debate^{5–8}. The saturating relationship suggests that some species are redundant with respect to a single function. However, short-term studies only address the effect of biodiversity on ecosystem functioning at a given point in time and under a given set of conditions. Some experimental results^{15–19} suggest that biodiversity can sometimes have an insurance value²⁰ by buffering ecosystem-level processes in a way analogous to that in which diverse investment portfolios spread financial risk and improve average performance in the longer term^{21–23}. Nevertheless, all of the research to date considers ecosystem processes examined individually, despite the fact that most ecosystems are managed or valued for several ecosystem services or processes: so-called ecosystem ‘multifunctionality’²⁴. If it is the case that a single species, or group of species, controls ecosystem functioning, then the remaining species are functionally redundant. Although it seems unlikely that a single species could control all ecosystem processes, it is possible that a single group of species may. However, if there is appreciable lack of overlap in the groups of species that influence different ecosystem processes, then higher levels of biodiversity will be required to maintain overall ecosystem functioning than indicated by analyses focusing on individual ecosystem processes in isolation.

To address the question of ecosystem multifunctionality, we used published data^{10,11} on seven ecosystem processes and properties (for brevity, hereafter processes) measured by the BIODEPTH project. BIODEPTH comprised a consortium of eight co-ordinated biodiversity experiments that manipulated plant diversity at different European grassland sites and monitored the response of a variety of ecosystem processes⁹. The ecosystem variables used for this analysis were above- and below-ground net biomass production, the pool of nitrogen in above-ground vegetation, resource levels above- (light interception) and below-ground (soil mineral nitrogen) and decomposition of lignin and cellulose (see Methods).

In the first stage of this analysis, we identified sets of species that had effects on each ecosystem process, using an information theoretic approach. For each site, we performed a backward-deletion multiple regression that included contrasts for the presence/absence of each species, and used the Akaike Information Criterion (AIC) to identify the most parsimonious set of species that influenced each ecosystem process (our results were robust to the inclusion of pairwise interactions, the exclusion of data from monoculture plots, and the use of an alternative information criterion: see Methods and Supplementary Information). For each ecosystem process at each site, we then selected the subset of species with effects on processes that would usually be considered desirable from an ecosystem services perspective. These were species with positive effects for all functions, except for soil inorganic nitrogen and light availability at ground level, where negative effects are consistent with lower levels of unconsumed resources and therefore greater uptake and less potential for nitrate leaching. We then analysed the resulting species counts for differences between sites and processes.

There were some differences between sites and processes in the number of species, x , that affected functioning (Fig. 1), although analysis of the results of the AIC multiple regressions revealed that the differences between processes were not strong enough to be conventionally significant (Poisson analysis of deviance: $\chi^2_6 = 10.5$, $P = 0.10$). The average number of species affecting a single ecosystem process, \bar{x} , ranged from 3.2 to 6.6 species depending on experimental site (Table 1), although these differences between experiments were also not conventionally significant ($\chi^2_7 = 13.3$, $P = 0.07$). The total number of species with effects on one or more ecosystem process at a site ranged from 8 to 18 (Table 1). This variation between experiments probably arises from a mixture of biological and design differences: they may partly reflect variation in the size of the species pools at the different sites, as those with the highest numbers of species with effects on functioning were also those with the largest species pools (see Methods and Supplementary Information).

We next looked at the overlap in these sets of species at each site that had AIC-detectable effects in the desirable direction on the different ecosystem processes. To do this we calculated the overlap, o , in the sets of species influencing each pair of processes, using Sørensen’s index:

$$o = \frac{|E_i \cap E_j|}{0.5(|E_i| + |E_j|)} \quad (1)$$

where $|E_i|$ is the number of species contributing to process i and

¹Institute of Environmental Sciences, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland. [†]Present address: Department of Zoology, University of Oxford, Tinbergen Building, South Parks Road, Oxford OX1 3PS, UK.

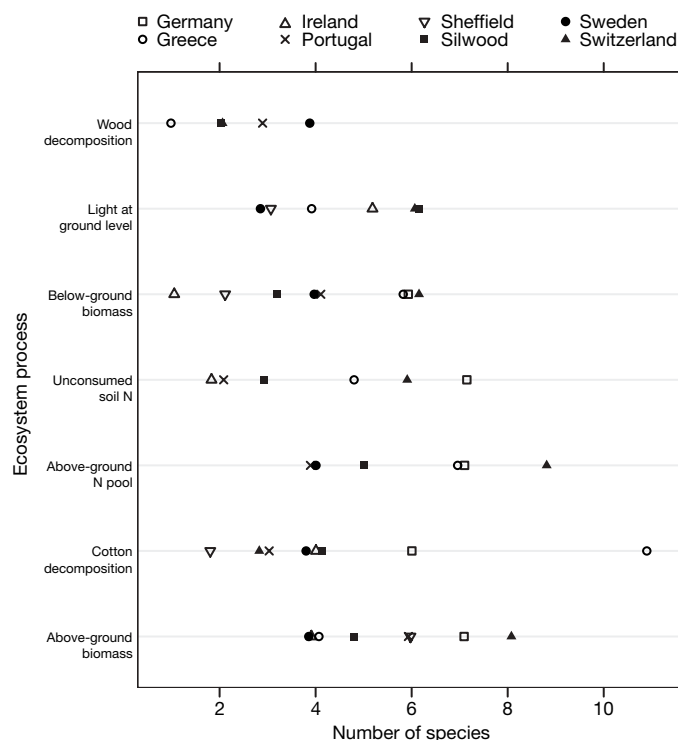


Figure 1 | Number of species with desirable effects on the suite of ecosystem processes measured in the different BIODEPTH project experiments. The number of species was identified by the AIC-based multiple regression (and species with effects with undesirable signs were then excluded).

$|E_i \cap E_j|$ is the number of species that contribute to both processes i and j . The average proportional overlap between the sets of species influencing a pair of processes, \bar{o} , ranged from 0.19 to 0.49 (Table 1) although these differences in \bar{o} between sites were not significant (quasi-maximum likelihood binomial analysis of deviance: $F_{7,111} = 1.08$, $P = 0.38$). We identified the single most important species affecting each process from the multiple regression minimally adequate model estimates, and calculated a simple measure of turnover as the proportion of ecosystem processes with different most-important species (the number of processes divided by the number of different species; where the maximum possible value of 1 indicates that the identity of the most important species differs for each process). Turnover in the most important species was generally high, ranging from 0.67 to 1, with an average of 0.78 (Table 1).

In the second stage of the analysis, we took the multiple regression results and examined how the number of species affecting functioning changed as more processes were considered. We took the subsets of species that had detectable and desirable effects on the ecosystem process (or combination of processes) in question at each site, and calculated how many species affected functioning for every possible combination of ecosystem processes (that is, all pair-wise combinations, each combination of three, and so on). Figure 2 plots all possible combinations of

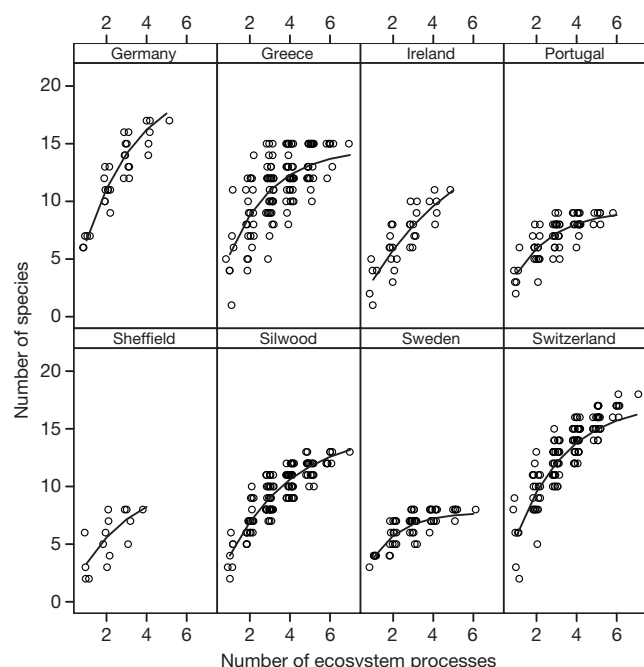


Figure 2 | Positive relationship between the range of ecosystem processes considered and the number of species that affect one or more aspect of ecosystem functioning. The points (jittered for clarity) show numbers of species required for all possible combinations of ecosystem processes. Lines are theoretical predictions from the model based on the average number of species required for a single process, \bar{x} , and the average overlap in the sets of species required for each pair of processes, \bar{o} , using equation (2).

ecosystem processes, to explore the complete range of behaviour observed in our experiments. At all sites, the envelope of points forms a concave-up curved wedge, with the apex determined by the number of species required for all functions, and the shortest side of the wedge given by the range of values observed for single ecosystem processes (Table 1). We also predicted how many species on average affect ecosystem functioning as numbers of processes increase, based on only the mean number of species with effects on a single ecosystem process, \bar{x} , and the mean average pairwise overlap, \bar{o} . The average number of species, S_E , required for E processes was predicted as (see Methods for a derivation):

$$S_E = \sum_{i=1}^E \binom{E}{i} \bar{x}(-\bar{o})^{(i-1)} \quad (2)$$

Predictions of S_E for each site are plotted as curves in Fig. 2. These curves are the average predictions given the simplifying assumptions made by using \bar{x} and \bar{o} . The spread of the points around the curves reflects differences in x for different processes and the variation in o for different pairs of processes. Two null hypotheses illustrate some extreme cases within the range of possible behaviour. At one extreme, if each species influences only a single process ($\bar{o} = 0$), and each process is only affected by a single species, it would generate a positive

Table 1 | Summary of ecosystem multifunctionality analysis

Site	Number of processes	Number of species	Number of species for all processes	Turnover	Mean number of species per process, \bar{x}	Mean overlap between process pairs, \bar{o}
Germany	5	31 (24)	17	0.80	6.6 (4.6–9.1)	0.32 (0.19–0.47)
Greece	7	23 (19)	15	0.57	5.4 (3.9–7.3)	0.37 (0.26–0.48)
Ireland	5	12 (11)	11	1.00	3.2 (1.9–5.0)	0.19 (0.06–0.39)
Portugal	6	14 (11)	9	0.67	3.7 (2.3–5.4)	0.40 (0.25–0.57)
Sheffield	4	12 (12)	8	1.00	3.3 (1.8–5.3)	0.31 (0.10–0.59)
Silwood	7	34 (24)	13	0.86	4.0 (2.7–5.7)	0.24 (0.17–0.40)
Sweden	6	12 (12)	8	0.67	3.8 (2.5–5.6)	0.49 (0.33–0.65)
Switzerland	7	48 (33)	18	0.71	5.7 (4.1–7.7)	0.33 (0.23–0.44)

'Number of processes' is the number of ecosystem processes measured per site. 'Number of species' is the number of species originally sown; shown in parentheses is the number present in biomass samples in year three of the experiment. 'Number of species for all processes' is the total (cumulative) number of unique species with AIC-detectable effects on at least one ecosystem process. 'Turnover' is the proportion of ecosystem processes with different most important species, as explained in the main text. Definitions of \bar{x} and \bar{o} are given in the main text; numbers in parentheses are the 95% confidence intervals.

linear relationship with a slope of $\bar{x} = 1$ (one species to one process). On the other hand, if all ecosystem processes are influenced by the same subset of species ($\bar{\sigma} = 1$), the slope would be zero with the intercept indicating the size of the group of species that affects ecosystem functioning. All observed relationships lie between these two extremes, showing a positive but saturating relationship between increasing numbers of ecosystem processes and the total number of species that influence ecosystem functioning (Fig. 2; Table 1).

Our study of ecosystem multifunctionality has limitations that should be addressed by future research. For example, the types of ecosystem stocks and flows analysed by the first generation of research on biodiversity and ecosystem functioning can be seen in some ways as the more-accessible ones. It would be interesting to repeat a similar analysis for an ecosystem where attempts had been made to identify and measure all of the ecosystem services considered important from a management perspective, and where the effects of all species could be precisely estimated with a well-replicated balanced design. Our study also manipulates only plant diversity, and our analysis therefore almost certainly underestimates levels of diversity needed for ecosystem multifunctionality. Our results will reflect the effects of other groups of organisms on ecosystem processes to the degree that the manipulation of plant species had knock-on effects to other groups, such as bacteria and mycorrhiza, but our study does not directly address the functional role played by these and other important groups. Better assessments of ecosystem multifunctionality could be made by manipulating multiple trophic levels simultaneously, and especially by studies that include groups such as microbes that are known to drive many ecosystem processes.

To our knowledge, our study is the first to quantitatively address the effects of biodiversity on multiple ecosystem processes simultaneously. Analysis of single ecosystem processes in isolation generally reveals a positive but saturating relationship with increasing biodiversity. Our analysis of ecosystem multifunctionality reveals that different processes are not affected by exactly the same species. Because different species affect different processes, maintaining multifunctional ecosystems will require greater biodiversity than suggested by studies focusing on single ecosystem processes in isolation.

METHODS SUMMARY

Ecosystem processes. The ecosystem process data used in our analysis are available online (<http://www.esapubs.org/archive/mono/M075/001/suppl-1.htm>) from Ecological Archives^{10,11} (see Methods).

Identifying species effects on ecosystem processes. At each site, we identified the most parsimonious set of species influencing each ecosystem function using a backward-elimination multiple regression analysis to identify a minimally adequate model²⁵ based on AIC. We used the stepAIC function in the MASS library²⁶ implemented in R 2.4.1²⁷. Models including pairwise interactions between species or excluding data from monocultures produced qualitatively similar results (Supplementary Information). We used an information-theoretic approach rather than one based on probability because information criteria have several advantages for the type of complex multiple regression analysis performed here²⁸ (Supplementary Information). Many alternative information criteria exist^{29,30}. We present results for AIC, as it is reputed to have the most solid foundation in theory and the best analytical behaviour²⁸, but analysis using the Bayesian Information Criteria (BIC) produced qualitatively similar results (Supplementary Information).

Ecosystem multifunctionality analysis. The mean number of species required per ecosystem process, \bar{x} , was estimated for each site using a Poisson generalized linear model analysing the main effects of site and process. The mean proportional overlap between the sets of species influencing each pair of processes, $\bar{\sigma}$, was estimated for each site using a quasi-maximum likelihood version of the binomial generalized linear model to account for under- or overdispersion²⁵. The relationship between the number of processes and the average number of species with effects on functioning was then predicted using equation (2).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 13 April; accepted 15 May 2007.

1. Loreau, M. *et al.* Biodiversity and ecosystem functioning: current knowledge and future challenges. *Science* **294**, 804–809 (2001).

2. Kinzig, A., Tilman, D. & Pacala, S. (eds) *The Functional Consequences of Biodiversity: Empirical Progress and Theoretical Extensions* (Princeton Univ. Press, Princeton, 2002).
3. Loreau, M., Naeem, S. & Inchausti, P. (eds) *Biodiversity and Ecosystem Functioning: Synthesis and Perspectives* (Oxford Univ. Press, Oxford, 2002).
4. Hooper, D. U. *et al.* Effects of biodiversity on ecosystem functioning: a consensus of current knowledge and needs for future research. *Ecol. Monogr.* **75**, 3–36 (2005).
5. Srivastava, D. S. & Velland, M. Biodiversity-ecosystem function: is it relevant to conservation? *Annu. Rev. Ecol. Evol. System.* **36**, 267–294 (2005).
6. Balvanera, P. *et al.* Quantifying the evidence for biodiversity effects on ecosystem functioning and services. *Ecol. Lett.* **9**, 1146–1156 (2006).
7. Cardinale, B. J. *et al.* Effects of biodiversity on the functioning of trophic groups and ecosystems. *Nature* **443**, 989–992 (2006).
8. Worm, B. *et al.* Impacts of biodiversity loss on ocean ecosystem services. *Science* **314**, 787–790 (2006).
9. Hector, A. *et al.* Plant diversity and productivity experiments in European grasslands. *Science* **286**, 1123–1127 (1999).
10. Hector, A. *et al.* Ecosystem effects of the manipulation of plant diversity in European grasslands: data from the BIODEPTH project. *Ecol. Archives* M075–001, S1 (2005); (<http://esapubs.org/Archive/mono/M075/001/suppl-1.htm>).
11. Spehn, E. M. *et al.* Ecosystem effects of the manipulation of plant diversity in European grasslands. *Ecol. Monogr.* **75**, 37–63 (2005).
12. Carlander, K. D. The standing crop of fish in lakes. *J. Fish. Res. Bd Can.* **12**, 543–570 (1955).
13. Darwin, C. & Wallace, A. On the tendency of species to form varieties; and on the perpetuation of varieties and species by natural means of selection *Linn. Soc. J.* **3**, 45–62 (1858).
14. Hector, A. & Hooper, R. E. Darwin and the first ecological experiment. *Science* **295**, 639–640 (2002).
15. McCann, K. S. The diversity and stability of ecosystems. *Nature* **405**, 228–233 (2000).
16. Cottingham, K. L., Brown, B. L. & Lennon, J. T. Biodiversity may regulate the temporal variability of ecological systems. *Ecol. Lett.* **4**, 72–85 (2001).
17. Loreau, M. *et al.* in *Biodiversity and Ecosystem Functioning: Synthesis and Perspectives* (eds Loreau, M., Naeem, S. & Inchausti, P.) 79–91 (Oxford Univ. Press, Oxford, 2002).
18. Steiner, C. F., Long, Z. T., Krumins, J. A. & Morin, P. Population and community resilience in multitrophic communities. *Ecology* **87**, 996–1007 (2006).
19. Tilman, D., Reich, P. B. & Knops, J. M. H. Biodiversity and ecosystem stability in a decade-long grassland experiment. *Nature* **441**, 629–632 (2006).
20. Yachi, S. & Loreau, M. Biodiversity and ecosystem productivity in a fluctuating environment: The insurance hypothesis *Proc. Natl Acad. Sci. USA* **96**, 1463–1468 (1999).
21. Doak, D. F. *et al.* The statistical inevitability of stability-diversity relationships in community ecology. *Am. Nat.* **151**, 264–276 (1998).
22. Tilman, D., Lehman, C. L. & Bristow, C. E. Diversity-stability relationships: Statistical inevitability or ecological consequence? *Am. Nat.* **151**, 277–282 (1998).
23. Lehman, C. L. & Tilman, D. Biodiversity, stability, and productivity in competitive communities. *Am. Nat.* **156**, 534–552 (2000).
24. Sanderson, M. A. *et al.* Plant species diversity and management of temperate forage and grazing land ecosystems. *Crop Sci.* **44**, 1132–1144 (2004).
25. McCullagh, P. & Nelder, J. A. *Generalized Linear Models* (Chapman and Hall, London, 1989).
26. Venables, W. N. & Ripley, B. D. *Modern Applied Statistics with S* (Springer, Berlin, 2002).
27. R Development Core Team. *A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, Austria, 2007).
28. Burnham, K. P. & Anderson, D. R. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach* (Springer, New York, 2002).
29. Ripley, B. D. in *Methods and Models in Statistics: In Honour of Professor John Nelder FRS* (eds Adams, N., Crowder, M., Hand, D. J. & Stephens, D.) 155–170 (Imperial College Press, London, 2004).
30. Taper, M. L. in *The Nature of Scientific Evidence* (eds Taper, M. L. & Lele, S. R.) 488–501 (Univ. Chicago Press, Chicago, 2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the BIODEPTH project members for releasing this data on Ecological Archives, E. Bazeley-White and L. Wacker for data management, M. Weilenmann for help with the manuscript, L. Jackson and the participants of the DIVERSITAS-BioMERGE 2006 workshop for discussion; W. Venables for advice on stepAIC; and M. Loreau, B. Schmid and L. Turnbull for comments.

Author Contributions The two-stage analysis of multifunctionality was conceived by A.H., the mathematical formulation by R.B., and the analysis and writing was performed jointly.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.H. (ahector@uwinst.uzh.ch).

METHODS

The BIODDEPTH experiments. The BIODDEPTH experiments were primarily designed to look at the influence of species and functional group richness on ecosystem processes. Random selection of species to form these gradients resulted in unequal replication for different species meaning that their effects are estimated with differing precision. The size of the species pool at each site reflected variation in background levels of diversity at the different grassland sites that provided the models for the experimental systems, and the highest level of diversity used in the species richness gradients was intended to approximate average levels of diversity in these surrounding natural grasslands³¹. Experiments at sites with larger pools tended to have a smaller proportion of species that were present in many plots and a larger proportion with lower replication because the experiments were of similar size regardless of the differences in the sizes of the species pool. The analyses presented in the main text use the whole diversity gradients, including monocultures, but the results are robust to their exclusion (see below).

Ecosystem processes. The data for the seven ecosystem processes used in our analysis are from year three of the BIODDEPTH project, and comprise measurements of: (1) above-ground (shoot) biomass; (2) below-ground (root) biomass; (3) total nitrogen pools in above-ground vegetation; (4) soil mineral nitrogen (nitrate + ammonium); (5) percentage of transmitted photosynthetically active radiation (PAR) at ground level; (6) decomposition (per cent per day) of wooden sticks (lignin); and (7) decomposition (per cent per day) of cotton strips (cellulose).

Identifying sets of species that affect ecosystem processes. Before model fitting, the response variables were transformed as indicated by the Box-Cox function in the MASS library for S-Plus and R²⁶. All ecosystem process variables (for brevity we use ecosystem processes to refer to both stocks and flows) were natural log transformed, except above-ground biomass (cube root) and soil inorganic N (untransformed). We performed single degree of freedom contrasts for the main effects of individual species at a given site within a backward-elimination multiple regression analysis, based on minimization of AIC, using the stepAIC function in the MASS library²⁶ implemented in R 2.4.1²⁷. Starting with a full model that contained a main effect for the presence/absence of every species at a given site, each species was removed from the model in turn and the AIC calculated. The AIC values of the resulting models were compared, and the species whose exclusion led to the greatest improvement (reduction) in the AIC value was permanently excluded. This was repeated until dropping any of the remaining species increased the AIC, at which point a minimally adequate model²⁵ had been selected based on AIC which identified the most parsimonious set of species influencing the ecosystem process in question. Models that included pairwise interactions between species or which excluded data from monocultures produced qualitatively similar general results (Supplementary Information). The AIC is defined as:

$$AIC = -2\ln(L) + 2k$$

where L is the model maximum likelihood (and $-2\ln(L)$ the deviance) and k is the number of model parameters. Schwartz's Bayesian Information Criterion (BIC), which has a harsher penalty for complexity and therefore tends to select simpler models, is defined as:

$$BIC = -2\ln(L) + \ln(n)k$$

where n is the sample size. We chose to repeat the analysis with the BIC because it is probably the most widely used alternative to AIC, and because the BIC and AIC come from different classes of information criteria (the dimensional-consistent and efficient classes, respectively).

Ecosystem multifunctionality analysis. The mean number of species required per ecosystem process, \bar{x} (with 95% confidence intervals), was estimated for each site using a Poisson generalized linear model analysing the main effects of site and process. The mean proportional overlap between the sets of species influencing each pair of processes, \bar{o} (and 95% confidence intervals), was estimated for each site using a quasi-maximum-likelihood version of the binomial generalized linear model to account for under- or overdispersion²⁵. Once \bar{x} and \bar{o} had been estimated, the relationship between the number of processes and the average number of species with effects on functioning was predicted using equation (2).

Predicting the number of species required for a given number of ecosystem processes. We estimated the expected number of species required for a given number of ecosystem processes using set theory. Given E sets (each comprising the species providing a given ecosystem process), the number of elements of the union of the sets can be defined according to the inclusion-exclusion principle³² as:

$$|A_1 \cup A_2 \cup \dots \cup A_E| = \sum_{1 \leq i \leq E} |A_i| - \sum_{1 \leq i_1 \leq i_2 \leq E} |A_{i_1} \cap A_{i_2}| + \sum_{1 \leq i_1 \leq i_2 \leq i_3 \leq E} |A_{i_1} \cap A_{i_2} \cap A_{i_3}| - \dots + (-1)^{(E-1)} |A_1 \cap A_2 \cap \dots \cap A_E| \quad (3)$$

We simplified equation (3) by making two assumptions. First, that all processes were controlled by the same number of species (that is, $|A_i| = x$, for all i where x is a constant). Second, that the overlap between the species contributing to pairs of processes, o , was also the same for all species pairs (that is, $|A_i \cap A_j| = o$, for all i and j , where o is a constant). Under these assumptions, and setting $S_E = |A_1 \cup A_2 \cup \dots \cup A_E|$, equation (3) simplifies to:

$$S_E = \binom{E}{1}x - \binom{E}{2}x \cdot o + \binom{E}{3}x \cdot o^2 - \dots + (-1)^{(E-1)} \binom{E}{E}x \cdot o^{(E-1)} \quad (4)$$

For each site, we estimated the constants x and o from the data as the mean number of species per ecosystem process (\bar{x}) and the mean proportion of species shared between pairs of processes (\bar{o}), respectively. Simplifying equation (4) and replacing x and o with \bar{x} and \bar{o} results in equation (2) in the main text. Note that equation (2) in the main text can be simplified algebraically to:

$$S_E = \frac{\bar{x} \cdot (1 - (1 - \bar{o})^E)}{\bar{o}} \quad (5)$$

This form can be used to infer the number of species that would theoretically provide all ecosystem processes by setting E to ∞ :

$$S_{\max} = \lim_{E \rightarrow \infty} \frac{\bar{x} \cdot (1 - (1 - \bar{o})^E)}{\bar{o}} = \frac{\bar{x}}{\bar{o}} \quad (6)$$

31. Hector, A. *et al.* Biodiversity and ecosystem functioning: reconciling the results of experimental and observational studies. *Funct. Ecol.* (in the press).

32. Comtet, L. *Advanced Combinatorics: The Art of Finite and Infinite Expansions* 176–177 (Reidel, Dordrecht, 1974).

Derivation of pluripotent epiblast stem cells from mammalian embryos

I. Gabrielle M. Brons¹, Lucy E. Smithers², Matthew W. B. Trotter², Peter Rugg-Gunn^{1†}, Bowen Sun¹, Susana M. Chuva de Sousa Lopes³, Sarah K. Howlett⁴, Amanda Clarkson⁵, Lars Ahrlund-Richter⁶, Roger A. Pedersen¹ & Ludovic Vallier¹

Although the first mouse embryonic stem (ES) cell lines were derived 25 years ago^{1,2} using feeder-layer-based blastocyst cultures, subsequent efforts to extend the approach to other mammals, including both laboratory and domestic species, have been relatively unsuccessful. The most notable exceptions were the derivation of non-human primate ES cell lines³ followed shortly thereafter by their derivation of human ES cells⁴. Despite the apparent common origin and the similar pluripotency of mouse and human embryonic stem cells, recent studies have revealed that they use different signalling pathways to maintain their pluripotent status. Mouse ES cells depend on leukaemia inhibitory factor and bone morphogenetic protein, whereas their human counterparts rely on activin (INHBA)/nodal (NODAL) and fibroblast growth factor (FGF). Here we show that pluripotent stem cells can be derived from the late epiblast layer of post-implantation mouse and rat embryos using chemically defined, activin-containing culture medium that is sufficient for long-term maintenance of human embryonic stem cells. Our results demonstrate that activin/Nodal signalling has an evolutionarily conserved role in the derivation and the maintenance of pluripotency in these novel stem cells. Epiblast stem cells provide a valuable experimental system for determining whether distinctions between mouse and human embryonic stem cells reflect species differences or diverse temporal origins.

We initially determined that prolonged culture of human ES cells in chemically defined medium⁵ (CDM) containing activin A and FGF2 (CDM/AF) maintained their fundamental characteristics (See Supplementary Data). We then tested similar conditions for derivation of pluripotent cells from pre- and post-implantation rodent embryos. Isolated inner cell masses (ICM) grown in CDM/AF never gave rise to pluripotent cell lines, but underwent rapid differentiation (Table 1 and Supplementary Fig. 2). In contrast, when

late epiblast layers were dissected from pre-gastrulation stages (5.75 days post coitum (d.p.c.)/embryonic day (E)5.75) for a *B6* × *CBA* F₁ genetic background or 6.5 d.p.c. for a *NOD* genetic background) and cultured for 24 h in CDM/AF, they formed colonies of compact cells with high nucleo-cytoplasmic ratios (Fig. 1c), a characteristic trait of pluripotent stem cells. Cells on the periphery started to differentiate the following day, producing a broad ring of stromal cells surrounding smaller aggregates of compact cells, which grew to form larger colonies of compact cells 4–5 days later (Fig. 1a–j). Such colonies were picked, fragmented into smaller clumps using collagenase and mechanical dissociation, and similarly passaged at 5-day intervals. Immunostaining after 20 passages showed that these colonies of compact cells expressed the pluripotency markers Oct-4 (*Pou5f1*), Nanog and SSEA-1 (*Fut4*). They were designated epiblast stem cells (EpiSCs) on the basis of their origin from pure late epiblast cell layers. The high rate of derivation (83% for *B6* × *CBA* F₁ background and 93% for the *NOD* genetic background; Table 1) was striking for strains considered ‘non-permissive’ for mouse ES cell derivation⁶. Expression of Oct-4, Nanog and SSEA-1 approached 95% and was maintained for more than 40 passages (Fig. 1k; Supplementary Fig. 3), suggesting that EpiSCs represent a near-homogenous population of pluripotent cells.

EpiSCs grew as flat, compact colonies and were thus morphologically distinct from mouse ES cells, which form rounded colonies (Fig. 1k). Unlike mouse ES cells, passaging of EpiSCs using trypsin or other single-cell dissociation methods induced widespread cell death. No EpiSC lines could be derived in the presence of LIF and/or BMP4 (Supplementary Fig. 4a), the two factors required for mouse ES cell derivation and self-renewal. On the other hand, the activin receptor inhibitor SB431542 induced rapid differentiation of EpiSCs (Supplementary Fig. 3 and Supplementary Fig. 4b), showing that EpiSC pluripotency depends strictly on activin/Nodal signalling. Finally, mouse

Table 1 | Efficiency of derivation of EpiSCs from mouse and rat embryos

Number of epiblasts		d.p.c.	Temperature	Matrix	Growth factors added to CDM	Lines derived		Rate (%)
Mouse	Rat					Mouse	Rat	
<i>B6</i> × <i>CBA</i>	<i>NOD</i>					<i>B6</i> × <i>CBA</i>	<i>NOD</i>	
Wistar	Sprague-Dawley						Wistar	Sprague-Dawley
32		3.5	37 °C	Fibro	Activin + FGF2	0		0
6		5.75	37 °C	Fibro	Activin + FGF2	5		83
	9	6.5	37 °C	Fibro	Activin + FGF2		9	100
	5	6.5	37 °C	Fibro	Activin + FGF2 + noggin		4	80
	14	7.75	38 °C	FCS	Activin		4	28.5
	20	7.75	37 °C	Fibro	Activin			1
								5

¹Department of Surgery and Cambridge Institute for Medical Research, Addenbrooke's Hospital, University of Cambridge, Cambridge CB2 0XY, UK. ²CR-UK Viral Oncology Group, Wolfson Institute for Biomedical Research, UCL, Cruciform Building Gower Street, London WC1E 6BT, UK. ³Wellcome Trust/Cancer Research UK Gurdon Institute of Cancer and Developmental Biology and Department of Physiology, University of Cambridge, Tennis Court Road, Cambridge CB2, 1QR, UK. ⁴Juvenile Diabetes Research Foundation/Wellcome Trust Diabetes and Inflammation Laboratory, University of Cambridge, Cambridge CB2 0XY, UK. ⁵Medical Genetics Department Cambridge University Hospital NHS Foundation Trust, Kefford House Maris Lane Cambridge CB2 2FF, UK. ⁶Dept of Laboratory Medicine Clinical Research Centre, Karolinska University Hospital Karolinska Institutet 141 57 Stockholm, Sweden. †Present address: Hospital for Sick Children, Toronto Medical Discovery Tower 101 College Street, Toronto, Ontario M5G 1L7, TMDT, Canada.

EpiSCs could be derived in CDM supplemented with activin alone (data not shown), indicating that FGF was not strictly required during this process. However, FGF improved the overall quality of the cultures, suggesting that it reinforces the efficiency of activin signalling, as in human ES cells⁷. Taken together, these observations show that, despite their expression of standard markers of pluripotency, mouse EpiSCs differ significantly from mouse ES cells in their growth factor dependence, their colony morphology, and in being averse to passaging as single cells. Importantly, EpiSCs were also derived from rat embryos (which have not previously yielded ES cells through conventional derivation methods) at pre-gastrulation stages (E7.5–7.75) (Fig. 1a–j, Fig. 1k and Supplementary Information), suggesting that activin/Nodal signalling may have a broader role than that of LIF and BMP4 in maintenance of pluripotency in embryonic stem cells.

We characterized EpiSCs and their differentiated progeny using both quantitative (Q-PCR) and qualitative (PCR, immunostaining) methods. (Because of the apparent similarity of mouse and rat EpiSCs, we focused this further characterization on mouse EpiSCs, except where noted). EpiSCs expressed higher levels of Oct-4 and Nanog and similar levels of Sox2 as compared with mouse ES cells (Fig. 2a). Conversely, EpiSCs did not express Rex1 (Zfp42), a specific marker of the ICM⁸ that is silenced in early epiblast cells just after implantation but is expressed in mouse ES cells (Fig. 2a). EpiSC transcript levels for Gbx2, a marker of mouse ICM and ES cells, was one fourth that of mouse ES cells (Supplementary Fig. 5a). On

the other hand, transcripts coding for FGF5 and Nodal, two genes specifically expressed in the late epiblast layer⁸ after implantation, were clearly detected in EpiSCs, in contrast to their low levels in mouse ES cells (Fig. 2a). Interestingly, early primitive ectoderm-like (EPL) cells that were established by growing mouse ES cells in HepG2-conditioned medium in the absence of LIF⁹ also express FGF5, but not Rex1. However, EPL cells revert to ES cells in the presence of LIF—which is not possible with EpiSCs (data not shown)—and EPL cells are unable to differentiate *in vitro* into extra-embryonic tissues, which can be generated from EpiSCs grown in the presence of BMP4 (see below), thus excluding the possibility that EpiSCs are EPL cells. Curiously, alkaline phosphatase activity, which generally marks pluripotent cells (including human ES cells and mouse ES cells), was present in the late epiblast layer itself, but disappeared abruptly on culture and was undetectable in EpiSCs (Fig. 2b; Supplementary Fig. 5b). The lack of alkaline phosphatase activity also distinguishes EpiSCs from germ cells because the activity persists in primordial germ cells^{10,11} in the gastrulating embryo¹² and in their *in vitro* derivatives, the embryonic germ cells. In addition, EpiSCs did not express germ cell markers such as Blimp1 and Stella (Supplementary Information), thereby confirming that EpiSCs are not derivatives of primordial germ cells.

To define further the molecular properties of EpiSCs, we performed a global analysis of their expression profile using gene expression microarrays. Expression profiles of EpiSCs were compared to those

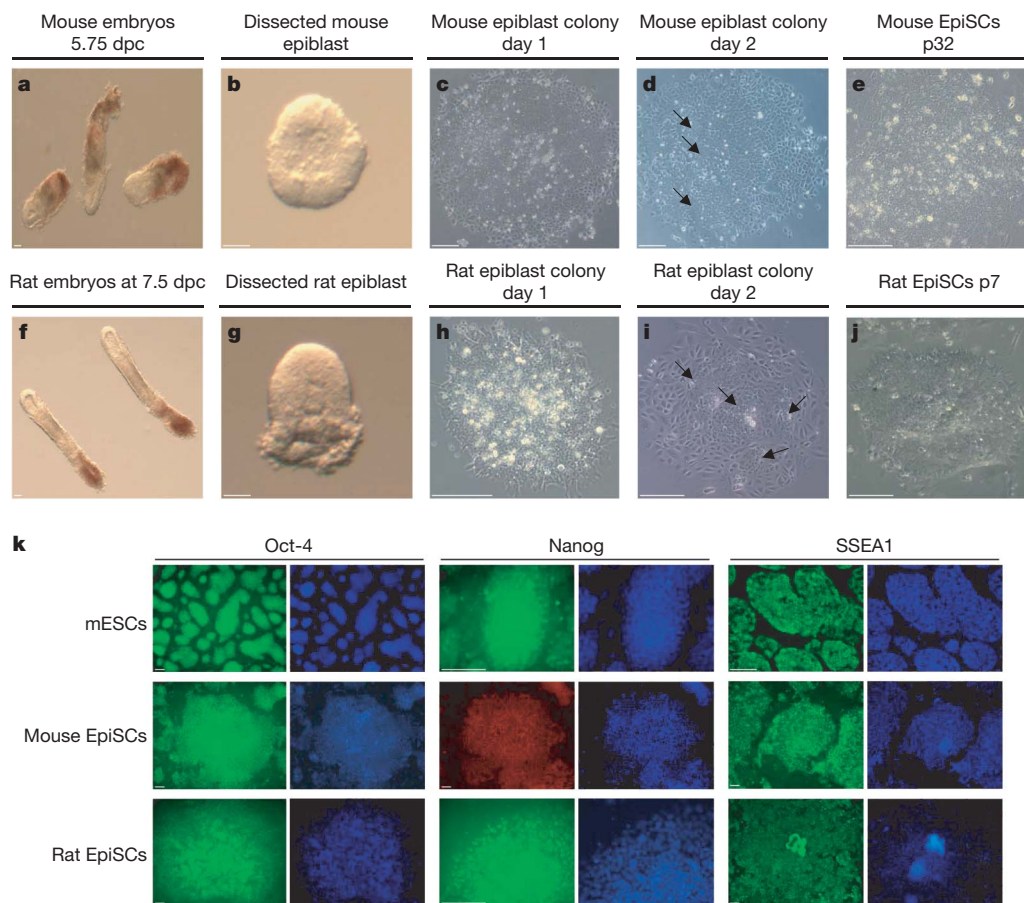


Figure 1 | Derivation of pluripotent epiblast stem cells (EpiSCs) from the late epiblast layer of embryos at post-implantation stages. a–j, Successive stages in EpiSC derivation from mouse (top) and rat (bottom) embryos. a, f, Mouse and rat embryos at pre-gastrulation stages (respectively, 5.75 d.p.c and 7.75 d.p.c). b, g, Pure epiblast layer obtained after removal of the extra-embryonic tissues. c, h, Epiblast layer grown for 24 h in CDM supplemented with activin (and FGF2 for mouse embryos). d, i, Epiblast outgrowth after 48 h of culture. Arrows indicate compact colonies of undifferentiated cells

surrounded by stromal cells. e, j, EpiSC colony after prolonged culture (32 passages or 9 months for mouse EpiSCs; 7 passages or 2 months for rat EpiSCs). Scale bar, 100 μ m. k, Expression of pluripotency markers in mouse ES cells, in mouse EpiSCs and in rat EpiSCs. Oct-4, Nanog and SSEA-1 expression were analysed by immunofluorescence in the mouse ES cell (mESC) line R1 of 129 strain at passage (p)35, in mouse EpiSCs of the NOD genetic background at p20, and in rat EpiSCs of the Wistar strain at p15. Scale bar, 100 μ m. Nuclei are shown by Hoechst staining.

of ICMs from mouse blastocysts, late epiblast cells from post-implantation mouse embryos, and mouse ES cells using multidimensional scaling (Fig. 2c). This showed that EpiSCs had fewer differences with dissected late epiblast than with ICM or mouse ES cells, suggesting that EpiSCs are transcriptionally similar to their embryonic tissue source. Despite sharing expression of core transcription factor circuitry with mouse ES cells, the EpiSCs established by culturing late epiblast layers in CDM/AF differ from mouse ES cells in other key attributes, suggesting that EpiSCs are a novel type of embryonic stem cell that closely resembles the late epiblast layer of the intact

post-implantation mouse embryo from which they have been derived.

To assess the extent of pluripotency of EpiSCs, we used a combination of *in vivo* and *in vitro* assessments involving ectopic transplantation, chimaera production, and embryoid body formation to examine their differentiative potential. EpiSC colonies injected into the testis capsule of immunodeficient mice were capable of forming teratomas containing a wide variety of tissues, including muscle, cartilage, neuronal rosettes, liver and gut (Fig. 3a–d), thereby reaffirming the multi-lineage pluripotency of EpiSCs. Next, their

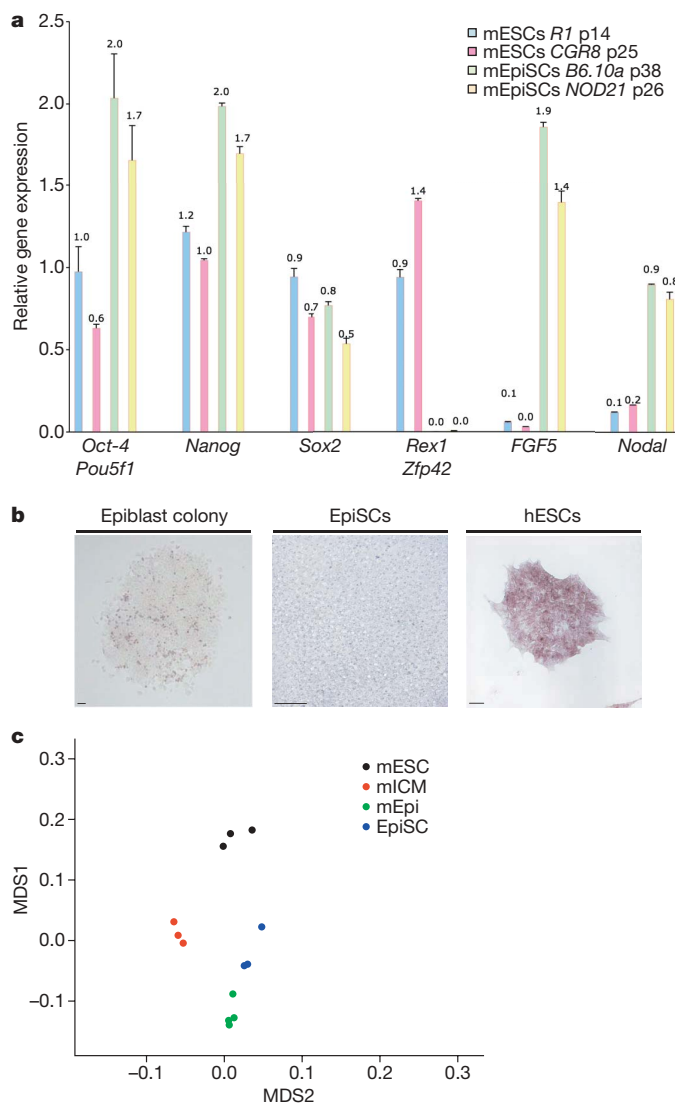


Figure 2 | Embryonic identity of EpiSCs. **a**, Expression of pluripotency markers specific for pre- and post-implantation embryonic stages in mouse ES cells and in mouse EpiSCs. Q-PCR analyses were performed to detect the genes denoted in two mouse ES cell lines (R1 p14 and CGR8 p25), and in 2 mouse EpiSC lines (B6.10a p38, NOD21 p26). Expression for each denoted gene was normalized with B2m (beta-2 microglobulin) and the resulting comparative levels, calculated from cycle thresholds, are indicated. Similar results were obtained in three independent experiments and error bars indicate their standard deviation. **b**, Absence of alkaline phosphatase activity in EpiSCs. Enzymatic activity for alkaline phosphatase was analysed in mouse epiblast cultured for 12 h after dissection and mouse EpiSCs of B6 × CBA F₁ genetic background (left and middle panels). Human (h)ESCs were used as positive controls (right panel). Scale bar, 100 μm. **c**, Comparison of expression profiles of ICM from embryos at the blastocyst stage, late epiblast from embryos at post-implantation stages, mouse ES cells and mouse EpiSCs. Microarray analysis of the transcriptional profile of ICM (mICM), epiblast (mEpi), mouse ES cells (mESCs) and mouse EpiSCs (EpiSC) were compared using an MDS plot (see Supplementary Information for full details of analysis).

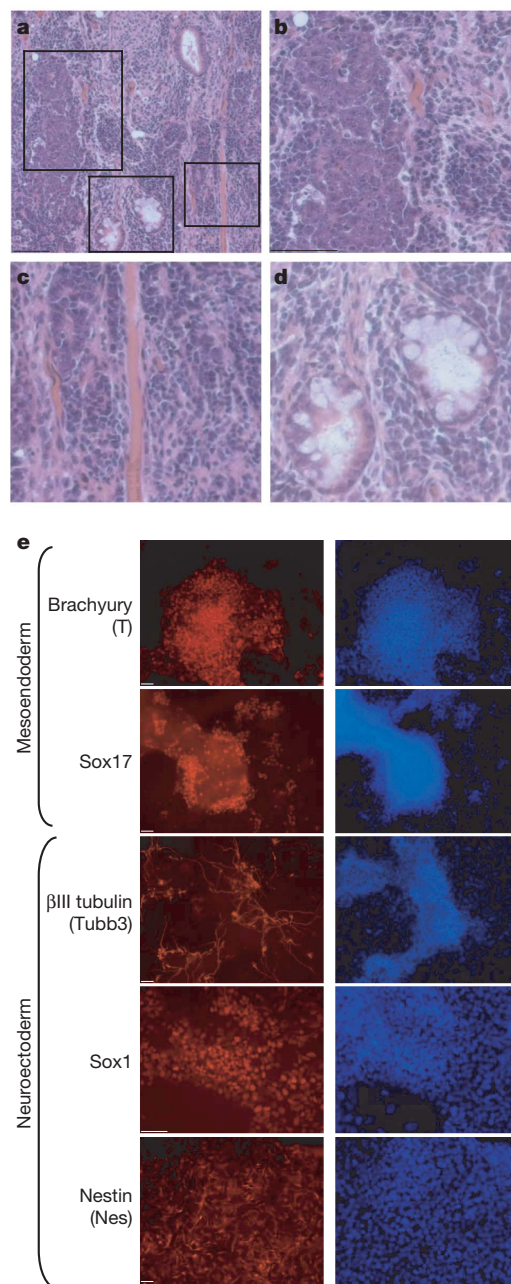


Figure 3 | EpiSCs are capable of differentiating into the three primary germ layers *in vitro* and *in vivo*. **a**, Teratomas from mouse EpiSCs. **a–d**, Differentiated derivatives of all three embryonic germ layers. Areas enclosed in boxes in **a** are enlarged in panels **b** (neuronal ganglion), **c** (striated muscle) and **d** (gut epithelium with Goblet cells). Scale bar, 100 μm. **e**, Expression of mesoderm and neuroectoderm markers in differentiated derivatives of EpiSCs. EpiSCs (NOD21 p20) were grown in culture conditions inducing neuroectoderm or mesendoderm differentiation. Expression of brachyury, Sox17, βIII tubulin, Sox1 and nestin was then analysed by immunostaining. Scale bar, 50 μm.

capacity for integration into pre-implantation-stage mouse embryos was examined by injecting EpiSCs into mouse blastocysts or by aggregating mouse embryos at the 8 cells or morula stage with EpiSC clumps (see Supplementary Information). Only two chimaeras were obtained out of 385 injected blastocysts, and germline transmission was not observed. Together these results suggest that, unlike the case for mouse ES cells, the pre-implantation embryo does not represent a compatible environment for EpiSCs, possibly owing to a developmental asynchrony that limits the ability of EpiSCs to colonise the host embryos. Alternatively, sparse chimaerism could be explained by a limited capacity of EpiSCs for development into early cell lineages.

Accordingly, to examine their differentiative capacity in greater detail, we determined whether EpiSCs could differentiate into a wide variety of cell types *in vitro*. Differentiation was achieved either by growing EpiSC colonies as embryoid bodies (Supplementary Information) or by differentiating them as monolayers into mesoderm, endoderm and ectoderm cells using a protocol devised for human ES cells (L. V., Touboul T., Chng, Z., Millan, E., Trotter, M., Weber, A. & R.A.P., unpublished observations). Generation of derivatives of the three germ layers was confirmed by the expression of early and late markers (Supplementary Fig. 7a) including brachyury (mesoderm), Sox17 (endoderm), Sox1, nestin and β -III tubulin (neuroectoderm) (Fig. 3e). As a further assessment of their *in vitro* differentiative capacity, EpiSCs were exposed to BMP4, which has been shown to induce differentiation of human ES cells into primitive endoderm and trophectoderm¹³. EpiSCs grown in CDM supplemented with BMP4 in the absence of activin signalling rapidly differentiated into cells expressing markers of primitive endoderm (Sox7, Gata4, Gata6) and trophectoderm (Cdx2, Hand1, Eomes, H19, α CG/Cga) (Supplementary Fig. 7b and 7c). Moreover, brachyury, Sox17 and Mixl1 were not detected in these conditions (Supplementary Fig. 7b), suggesting that BMP4 is insufficient to drive differentiation of EpiSCs into mesendoderm.

Finally, similar results were obtained with clonal sublines of mouse EpiSCs (Supplementary Information), demonstrating that single EpiSC cells were capable of recapitulating the entire range of differentiative outcomes and thereby ruling out the possibility of a mixed cell population with diverse capabilities. All together, these results demonstrate that EpiSCs were capable of differentiating into derivatives of all three primary germ layers and into extra-embryonic tissues *in vitro*. Such broad differentiative capacity reaffirms the pluripotent status of EpiSCs, thus leading to the conclusion that limited EpiSC contribution in chimaeras is attributable to developmental asynchrony. Taken together, these results demonstrate that pluripotent stem cells can be cultured from widely separated mammalian species using chemically defined culture medium supplemented with activin. On this basis, we conclude that the activin/Nodal pathway has a central role in pluripotency of a new type of embryonic stem cells derived from mammalian embryos, specifically representing the late epiblast cell population just before gastrulation.

EpiSCs differ from mouse ES cells not only in their embryonic tissue of origin but also in the signalling pathway maintaining their pluripotent status (activin/Nodal versus LIF). This dependence of EpiSC pluripotency on activin/Nodal signalling also highlights their resemblance to late epiblast *in vivo*. Indeed, recent publications have established that embryos mutant for Nodal lack the expression of the pluripotency markers Oct-4 and Nanog, instead ectopically expressing markers of anterior neuroectoderm^{14,15}. Our results extend these recent findings by showing that activin/Nodal signalling not only is necessary for blocking neuroectoderm specification *in vivo*, but also that it is sufficient for maintenance of pluripotency in stem cells cultured from the late epiblast. The dependency of mouse ES cells on LIF signalling has focused attention on the early epiblast (E4.5 or equivalent) as their likely tissue of origin¹⁶. This reflects the narrow temporal role of LIF/GP130 signalling, which is required for mouse ICM viability only during delayed implantation. Moreover, mouse

embryos that are homozygous mutants for Nodal¹⁷ or its type I and II receptors^{18,19}, progress well beyond the blastocyst stage, arresting only at pre-gastrula stages, consistent with the hypothesis that LIF and activin/Nodal signalling immortalize stem cell populations at distinct developmental stages, namely, early and late epiblast, respectively.

Intriguingly, EpiSCs share with human ES cells not only a dependence on activin/Nodal signalling but also other key differences from mouse ES cells that have previously been attributed to species divergence (flattened colony morphology, inefficient clonal propagation and a limited capacity for colonising pre-implantation embryos²⁰). Moreover, despite having similar core transcription factor circuitry, human ES cells and mouse ES cells have substantially different target genes for Oct-4 and Nanog^{21,22}. Also, the epigenetic stability of human ES cells seems to be greater than mouse ES cells, as shown here and previously by their extent of monoallelic imprinted gene expression^{23,24}. Finally, EpiSCs (like human ES cells¹³) have been shown to differentiate into trophectoderm in the presence of BMP4, whereas mouse ES cells have little or no capacity for contribution to either primitive endoderm or trophectoderm lineages in chimaeric embryos²⁵ and differentiate into trophectoderm only when their *Oct-4* gene is mutated by homologous recombination²⁶. In summary, EpiSCs are temporally defined, embryonic stem cells that add a strategically important experimental system for understanding the development of pluripotency. It has not escaped our attention that the unique properties of human ES cells, now seen to be shared with EpiSCs, could reflect a similar late epiblast origin. Understanding the relationships between cell-surface-mediated growth factor activities, core transcription factor circuitry and the epigenetic modifications accompanying pluripotency in EpiSCs could contribute significant insights to this issue.

METHODS

For derivation of EpiSCs, the late epiblast layer was dissected intact from pre-gastrula stage mouse [E5.75, (B6XCBA)F1 mated *inter se*; E6.5, NOD mated *inter se*] or rat [E7.5–7.75 Wistar or Sprague Dawley] embryos after 15–20 min treatment at 4°C in Cell Dissociation Buffer (Gibco). Dissection was accomplished in Calcium and Magnesium free Flushing and Handling Medium using glass needles to completely separate the epiblast layer from the extraembryonic tissues. Epiblast layers were then placed in chemically defined medium (CDM)⁵, supplemented with Activin A (20 ng/ml, RandD systems) and FGF2 (12 ng/ml, RandD systems). Every 4–5 days, cells were harvested using 5 mg/ml collagenase IV (Invitrogen) then plated into plates (Costar) pre-coated either with fetal bovine serum (FBS) or with 15 μ g/ml of human Fibronectin (Chemicon) for 20 min at 37°C and then washed twice in PBS. (Additional methods can be found in Supplementary Information).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 15 January; accepted 24 May 2007.

Published online 27 June 2007.

- Evans, M. J. & Kaufman, M. H. Establishment in culture of pluripotential cells from mouse embryos. *Nature* **292**, 154–156 (1981).
- Martin, G. R. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc. Natl Acad. Sci. USA* **78**, 7634–7638 (1981).
- Thomson, J. A. *et al.* Isolation of a primate embryonic stem cell line. *Proc. Natl Acad. Sci. USA* **92**, 7844–7848 (1995).
- Thomson, J. A. *et al.* Embryonic stem cell lines derived from human blastocysts. *Science* **282**, 1145–1147 (1998).
- Johansson, B. M. & Wiles, M. V. Evidence for involvement of activin A and bone morphogenetic protein 4 in mammalian mesoderm and hematopoietic development. *Mol. Cell. Biol.* **15**, 141–151 (1995).
- Brook, F. A. *et al.* The derivation of highly germline-competent embryonic stem cells containing NOD-derived genome. *Diabetes* **52**, 205–208 (2003).
- Vallier, L., Alexander, M. & Pedersen, R. A. Activin/Nodal and FGF pathways cooperate to maintain pluripotency of human embryonic stem cells. *J. Cell Sci.* **118**, 4495–4509 (2005).
- Pelton, T. A., Sharma, S., Schulz, T. C., Rathjen, J. & Rathjen, P. D. Transient pluripotent cell populations during primitive ectoderm formation: correlation of *in vivo* and *in vitro* pluripotent cell development. *J. Cell Sci.* **115**, 329–339 (2002).

9. Rathjen, J. *et al.* Formation of a primitive ectoderm like cell population, EPL cells, from ES cells in response to biologically derived factors. *J. Cell Sci.* **112**, 601–612 (1999).
10. Resnick, J. L., Bixler, L. S., Cheng, L. & Donovan, P. J. Long-term proliferation of mouse primordial germ cells in culture. *Nature* **359**, 550–551 (1992).
11. Matsui, Y., Zsebo, K. & Hogan, B. L. Derivation of pluripotential embryonic stem cells from murine primordial germ cells in culture. *Cell* **70**, 841–847 (1992).
12. De Felici, M. & McLaren, A. Isolation of mouse primordial germ cells. *Exp. Cell Res.* **142**, 476–482 (1982).
13. Xu, R. H. *et al.* BMP4 initiates human embryonic stem cell differentiation to trophoblast. *Nature Biotechnol.* **20**, 1261–1264 (2002).
14. Camus, A., Perea-Gomez, A., Moreau, A. & Collignon, J. Absence of Nodal signaling promotes precocious neural differentiation in the mouse embryo. *Dev. Biol.* **295**, 743–755 (2006).
15. Mesnard, D., Guzman-Ayala, M. & Constam, D. B. Nodal specifies embryonic visceral endoderm and sustains pluripotent cells in the epiblast before overt axial patterning. *Development* **133**, 2497–2505 (2006).
16. Nichols, J., Chambers, I., Taga, T. & Smith, A. Physiological rationale for responsiveness of mouse embryonic stem cells to gp130 cytokines. *Development* **128**, 2333–2339 (2001).
17. Conlon, F. L. *et al.* A primary requirement for nodal in the formation and maintenance of the primitive streak in the mouse. *Development* **120**, 1919–1928 (1994).
18. Song, J. *et al.* The type II activin receptors are essential for egg cylinder growth, gastrulation, and rostral head development in mice. *Dev. Biol.* **213**, 157–169 (1999).
19. Gu, Z. *et al.* The type I serine/threonine kinase receptor ActRIA (ALK2) is required for gastrulation of the mouse embryo. *Development* **126**, 2551–2561 (1999).
20. James, D., Noggle, S. A., Swigut, T. & Brivanlou, A. H. Contribution of human embryonic stem cells to mouse blastocysts. *Dev. Biol.* **295**, 90–102 (2006).
21. Boyer, L. A. *et al.* Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122**, 947–956 (2005).
22. Loh, Y. H. *et al.* The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nature Genet.* **38**, 431–440 (2006).
23. Rugg-Gunn, P. J., Ferguson-Smith, A. C. & Pedersen, R. A. Epigenetic status of human embryonic stem cells. *Nature Genet.* **37**, 585–587 (2005).
24. Sun, B. W. *et al.* Temporal and parental-specific expression of imprinted genes in a newly derived Chinese human embryonic stem cell line and embryoid bodies. *Hum. Mol. Genet.* **15**, 65–75 (2006).
25. Beddington, R. S. & Robertson, E. J. An assessment of the developmental potential of embryonic stem cells in the midgestation mouse embryo. *Development* **105**, 733–737 (1989).
26. Niwa, H., Miyazaki, J. & Smith, A. G. Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nature Genet.* **24**, 372–376 (2000).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. McLaren for her support. We thank A. Smith for CGR8 cells, A. Nagy for R1 cells, P. Andrew for the SSEA-1 antibody, L. Wicker for access to NOD mice, and S. Thiru for advice in the teratoma analysis. This work was supported by an MRC International Appointments Initiative grant (R.A.P.), MRC/Juvenile Diabetes Research Foundation Centre funding (R.A.P., I.G.M.B.), Remedi (I.G.M.B.), the Wellcome Trust Functional Genomics Initiative on Stem Cells (L.E.S., M.W.B.T.), the Addenbrookes National Institute for Health Research Biomedical Research Centre, and a Diabetes UK Career Development fellowship (L.V.). We dedicate this paper to the memory of our colleague Isabelle Bouhon.

Author Contributions L.V. conceived the experiment and did the molecular analysis; I.G.M.B. derived and cultured the EpiSCs; R.A.P. performed the epiblast dissections and together with I.G.M.B. carried out the aggregation, chimaera and clonal assays; L.E.S. and M.T. obtained the microarray data; P.R-G. performed the epigenetic analysis; B.S. performed the blastocyst immunosurgery and ICM cultures; S.M.C.d.S.L. did the Stella-GFP dissection and Blimp1 staining; S.K.H. provided PCR analysis of chimaeras; A.C. did karyotyping of human ES cell lines; L.A-R. carried out the teratoma work; L.V., R.A.P. and I.G.M.B. analysed the data and co-wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to L.V. (lv225@cam.ac.uk).

METHODS

EpiSC and human ES cell culture in feeder-free and serum-free conditions.

For feeder- and serum-free culture, human ES cells (H9 and H1; WiCell, Madison), hSF-6 (UCSF, San Francisco) and EpiSCs were grown in chemically defined medium (CDM)⁵, supplemented with activin (10 ng ml^{-1} for human ES cells; 20 ng ml^{-1} for mouse and rat EpiSCs; RandD systems) and FGF2 (12 ng ml^{-1} , RandD systems for human ESCs and mouse EpiSCs). The composition of CDM was 50% IMDM (Gibco) plus 50% F12 NUT-MIX (Gibco), supplemented with $7\text{ }\mu\text{g ml}^{-1}$ of insulin (Roche), $15\text{ }\mu\text{g ml}^{-1}$ of transferrin (Roche), $450\text{ }\mu\text{M}$ of monothioglycerol (Sigma) and 5 mg ml^{-1} bovine serum albumin fraction V (Europabioproducts). Every 4–5 days, cells were harvested using 5 mg ml^{-1} collagenase IV (Invitrogen) or Accutase (BioWest) and then plated into plates (Costar) pre-coated either with fetal bovine serum (FBS) or with $15\text{ }\mu\text{g ml}^{-1}$ of human Fibronectin (Chemicon) for 20 min at $37\text{ }^{\circ}\text{C}$ and then washed twice in PBS. EpiSCs and human ES cells were regularly frozen in a medium containing 90% Serum Replacer (Invitrogen) and 10% DMSO (Sigma).

ICM immuno-surgery and epiblast dissection. For derivation at the blastocyst stage, ICMs were isolated from 3.5 d.p.c. blastocysts obtained from $B6 \times CBA\text{ F}_1$ mice mated *inter se* using immuno-surgery and then placed in CDM/AF on fibronectin. For derivation at post-implantation stages, the late epiblast layer was dissected intact from pre-gastrula stage mouse (E5.75, $B6 \times CBA\text{ F}_1$ mated *inter se*; E6.5, NOD mated *inter se*) or rat (E7.5–7.75 Wistar or Sprague Dawley) embryos after 15–20 min treatment at $4\text{ }^{\circ}\text{C}$ in Cell Dissociation Buffer (Gibco). Dissection was accomplished in calcium- and magnesium-free Flushing and Handling Medium using glass needles to completely separate the epiblast layer from the extra-embryonic tissues. Epiblast layers were then placed in CDM supplemented with growth factors (activin (20 ng ml^{-1}) and FGF2 (12 ng ml^{-1})) for culture as above.

Subcloning of EpiSCs. EpiSC colonies were dissociated into single cells after 5 min of treatment in 0.05% Trypsin-EDTA (Gibco) at $37\text{ }^{\circ}\text{C}$. Individual cells were then seeded in three different culture conditions. Subcloning in CDM/AF on fibronectin never gave any EpiSC colonies, confirming that clonal propagation efficiency of EpiSCs in chemically defined medium is very low. Two other approaches based on feeder cells were then used to generate EpiSC sublines. The first approach consisted of seeding individual NOD background EpiSCs on green fluorescent protein (GFP)-expressing EpiSC colonies that had been irradiated to block their proliferation. Using this approach, the clonal efficiency of EpiSCs was 5%; three non-fluorescent colonies were picked and amplified to study their property of differentiation. Another method of subcloning consisted of seeding individual NOD background EpiSCs on a layer of irradiated mouse embryonic fibroblasts in CDM/AF. The clonal efficiency of EpiSCs in these culture conditions was $\sim 15\%$ with single cells seeded either by limiting dilution or using glass pipettes; 11 EpiSC sublines derived on mouse feeders were picked and amplified in CDM/AF to study their capacity to differentiate into the three germ layers and into extra-embryonic tissues.

LETTERS

New cell lines from mouse epiblast share defining features with human embryonic stem cells

Paul J. Tesar^{1,2*}, Josh G. Chenoweth^{1*}, Frances A. Brook², Timothy J. Davies², Edward P. Evans², David L. Mack³, Richard L. Gardner² & Ronald D. G. McKay¹

The application of human embryonic stem (ES) cells in medicine and biology has an inherent reliance on understanding the starting cell population. Human ES cells differ from mouse ES cells and the specific embryonic origin of both cell types is unclear. Previous work suggested that mouse ES cells could only be obtained from the embryo before implantation in the uterus^{1–5}. Here we show that cell lines can be derived from the epiblast, a tissue of the post-implantation embryo that generates the embryo proper. These cells, which we refer to as EpiSCs (post-implantation epiblast-derived stem cells), express transcription factors known to regulate pluripotency, maintain their genomic integrity, and robustly differentiate into the major somatic cell types as well as primordial germ cells. The EpiSC lines are distinct from mouse ES cells in their epigenetic state and the signals controlling their differentiation. Furthermore, EpiSC and human ES cells share patterns of gene expression and signalling responses that normally function in the epiblast. These results show that epiblast cells can be maintained as stable cell lines and interrogated to understand how pluripotent cells generate distinct fates during early development.

At implantation, the embryo is composed of the epiblast, a pluripotent derivative of the inner cell mass (ICM), and the extra-embryonic ectoderm and endoderm tissues, which are fated to

generate supporting structures such as the placenta and yolk sac⁶ (Supplementary Fig. 1). Identified tissues from early post-implantation (embryonic day (E)5.5) mouse embryos were individually explanted onto a feeder layer of irradiated mouse embryonic fibroblasts. Cells from the extra-embryonic ectoderm and endoderm generated distinct extra-embryonic fates (Supplementary Fig. 2). The epiblast, which generates all the cell types of the adult, efficiently grew as an epithelial colony when explanted and expressed transcription factors associated with pluripotency, such as Oct3/4 (also known as Pou5f1) and Nanog (Fig. 1a and Supplementary Fig. 2)⁷.

These primary epiblast explants could not initially be propagated with the protocol used to passage mouse ES cells but could be extensively passaged (>30) when a human ES cell procedure was used. Mouse ES cells typically grow in small, compact, domed colonies. EpiSC colonies were larger and grew as a monolayer similar to human ES cells, and continued to express genes associated with pluripotency (Fig. 1b and Supplementary Fig. 3). Gene expression was analysed by hybridization to probes that represent all the transcripts in the mouse genome. Cluster analysis showed that independently derived EpiSC lines were similar to each other and distinct from mouse ES cells (Fig. 1c). The pattern of gene expression in EpiSCs was consistent with that of the post-implantation epiblast and not the pre-implantation ICM (Supplementary Fig. 3)⁸.

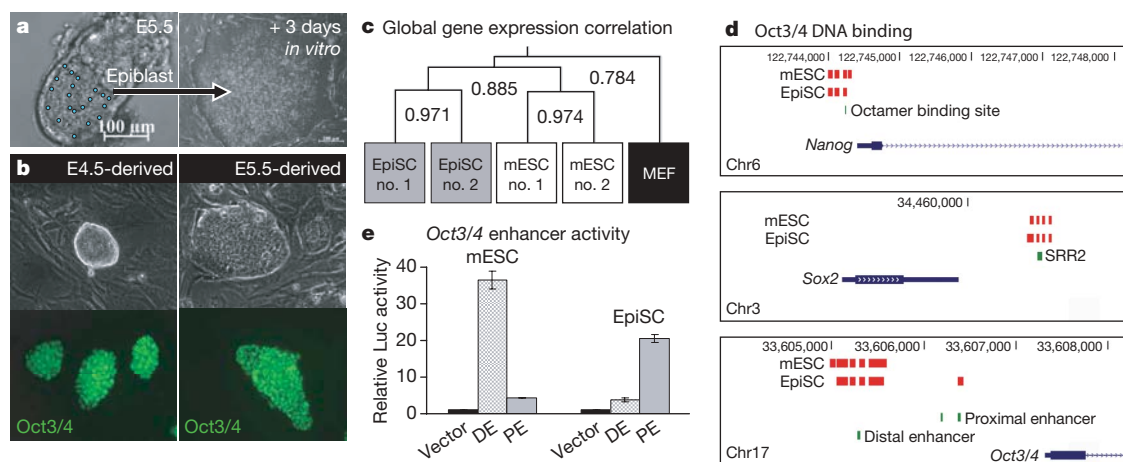


Figure 1 | Cell lines from the E5.5 murine epiblast can be isolated and maintained in culture. **a**, Epiblast tissue (outlined in blue) was dissected from the E5.5 mouse embryo and cultured for 3 days. **b**, Morphology and Oct3/4 expression of cell lines from pre- (left) and post-implantation (right) mouse embryos. **c**, Whole-genome cluster analysis of transcripts from two independent mouse ES cell (mESC) lines, two independent EpiSC lines, and

mouse embryonic fibroblasts. Correlation coefficients are indicated. **d**, Representation of Oct3/4 binding (shown in red) to regulatory regions of pluripotency genes *Nanog*, *Sox2* and *Oct3/4*. **e**, Quantification of *Oct3/4* distal enhancer (DE) and proximal enhancer (PE) reporter gene activity in mouse ES cells and EpiSCs. Data were obtained in triplicate and are presented as \pm s.d.

¹Laboratory of Molecular Biology, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, Maryland 20892, USA. ²Mammalian Development Laboratory, Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK. ³Stem Cell Biology Section, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, USA.

*These authors contributed equally to this work.

Chromatin immunoprecipitation combined with high density microarray (ChIP-on-chip) revealed that EpiSCs had key features of chromatin domains and transcription-factor-binding profiles that are restricted to pluripotent cells (Fig. 1d and Supplementary Fig. 4)^{9–12}. Two different regulatory sites, the distal and proximal enhancers, direct expression of the *Oct3/4* gene in the ICM and post-implantation epiblast respectively¹³. Transfection of reporter constructs showed that the distal and proximal enhancers have reciprocal abilities to direct gene expression in mouse ES cells and EpiSCs. The ICM-specific distal enhancer regulated expression in mouse ES cells and the epiblast-specific proximal enhancer promoted expression in the EpiSCs (Fig. 1d, e). This result shows that EpiSCs are distinct from mouse ES cells and capture transcriptional features of the post-implantation epiblast.

Chimaeric mice are efficiently generated when mouse ES cells are introduced into the pre-implantation embryo. The developmental potential of EpiSCs was tested by injection into E3.5 host blastocysts. As judged by coat and eye colour contribution, no chimaeras were found among 35 pups from 49 injected and transferred blastocysts in two different experiments using two separate cell lines. During the course of these injections, typical chimaerism levels were obtained with mouse ES cells (40 chimaeras from 59 pups). Because EpiSCs do not survive well as single cells, we examined their potential to contribute to embryogenesis through morula aggregation. When aggregated with morulae ($n = 23$) and cultured for 24 h, we saw no

indication that the EpiSCs could integrate effectively into the pre-implantation embryo (Fig. 2a). Strikingly, in the majority of cases, the cells impaired or delayed the development of the embryo. The EpiSCs either remained distinct from the embryo (Fig. 2a', note that the embryo is delayed relative to normal development of the unaggregated control in Fig. 2a) or were incorporated within the embryo but remained in the cavity of the blastocyst and not associated with ICM or trophectoderm (Fig. 2a''). Native epiblast cells of the post-implantation embryo also do not incorporate into the ICM but are clearly pluripotent^{14,15}. Owing to cells of epiblast origin being incompatible with the pre-implantation embryo environment, we tested the potency of our cells in two further ways.

Teratomas, disorganized tumours containing multiple differentiated cell types, are formed when pluripotent cells are grafted into ectopic sites. EpiSCs efficiently formed teratomas from all injections ($n = 12$) into athymic nude mice. Teratomas contained extensively differentiated cell types including ganglion cells, cartilage, muscle, adipocytes, keratinocytes, ciliated endodermal epithelium with secretory goblet cells, epithelial rosettes, and gut (Fig. 2b and Supplementary Fig. 5).

ES cells also differentiate effectively in culture when grown in suspension as aggregates. EpiSCs efficiently formed these embryoid bodies when grown in suspension, and culture of embryoid bodies in serum-free medium resulted in differentiation to predominantly neuroectodermal fates (Fig. 2c). Fetal bovine serum is known to promote mesodermal and endodermal differentiation¹⁶. When embryoid bodies were generated from EpiSCs in the presence of fetal bovine serum, nearly all embryoid body outgrowths contained contracting cardiomyocytes. Cardiomyocyte differentiation was further confirmed by Nkx2-5 immunoreactivity and expression of a cardiac-specific isoform of troponin T (Fig. 2c). Multinucleated, nebulin-positive skeletal myofibres were also present (Fig. 2c). Endoderm-like cells with a characteristic morphology and expressing the transcription factors *Gata6* and *Sox17* were also observed (Supplementary Fig. 6). The embryoid body differentiation strategy also generated endothelial cells with a typical granular staining for von Willebrand factor (Supplementary Fig. 5). Although the plating efficiency was low (~1–4%), single EpiSCs could be subcloned. The differentiation repertoire of the clonal cell lines was indistinguishable from the parental line (Supplementary Fig. 6). These data show that EpiSCs have a wide developmental potency.

To define better their developmental state, the pattern of gene expression in three independent lines of mouse ES cells and EpiSCs was assessed by hybridization to whole-genome expression arrays. The data show that *Oct3/4*, *Nanog* and *Sox2* were expressed at equivalent levels in mouse ES cells and EpiSCs (Fig. 3a). Transcripts from genes associated with the ICM such as *Pecam1*, *Tbx3* and *Gbx2* were expressed by mouse ES cells and significantly decreased or not detected in EpiSCs. Transcripts from genes associated with the epiblast and early germ layers such as *Otx2*, *Eomes*, *Foxa2*, *brachyury* (*T*), *Gata6*, *Sox17* and *Cer1* were expressed at higher levels in EpiSCs (Fig. 3a). Genes expressed in the epiblast and early germ layers were also detected in human ES cell cultures (shown in red in Fig. 3a), consistent with previous studies showing differences in gene expression between mouse and human ES cells^{17,18}.

To confirm these differences, we performed ChIP-on-chip analysis using custom arrays to detect histone modifications indicative of gene induction (H3K4me3) and repression (H3K27me3)⁷. The chromatin status around the transcriptional start sites was examined for the *Stella* (*Dppa3*), *Otx2* and *Nanog* genes in mouse ES cells, EpiSCs and human ES cells (red circles in Fig. 3b and Supplementary Fig. 7). Consistent with its restricted expression to mouse ES cells, *Stella* was only associated with inductive H3K4me3 in those cells, whereas *Nanog* was associated with this mark in all three cell types. *Otx2* had the inductive mark in all three cell types but also the repressive H3K27me3 mark in mouse ES cells. These data show that EpiSCs and human ES cells have common mechanisms of epigenetic regulation of transcription, which are distinct from mouse ES cells.

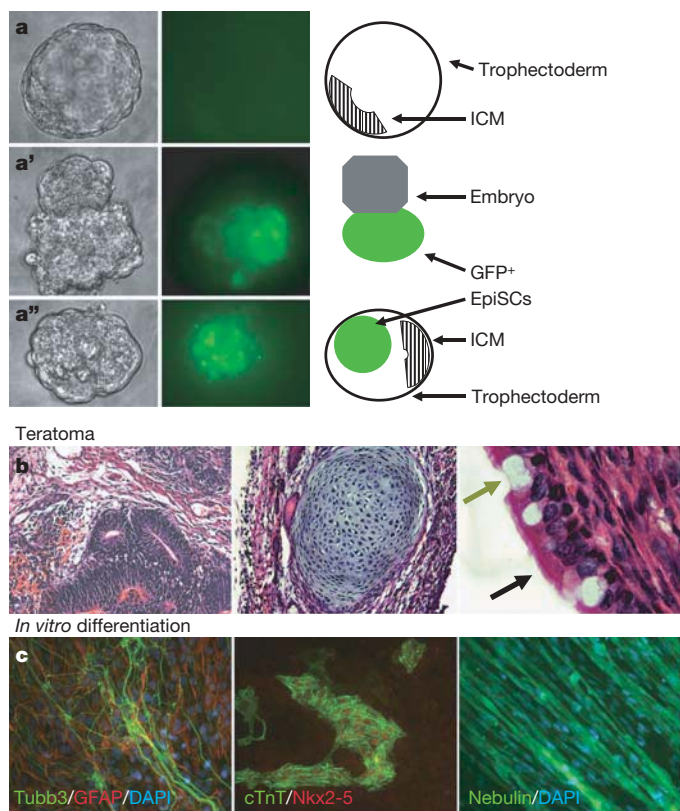


Figure 2 | EpiSC lines are pluripotent and differentiate *in vitro* and *in vivo*. **a**, EpiSCs do not efficiently incorporate into the pre-implantation embryo after aggregation with morulae. Shown are resulting structures after 1 day of culture from an unaggregated control morula (**a**) and green fluorescent protein (GFP)⁺ EpiSCs aggregated with individual morulae (**a'**, **a''**). Schematics are shown to the right to highlight aggregation results. **b**, Haematoxylin and eosin stained histological sections of EpiSC-derived teratomas. Shown are, from left to right, epithelial rosettes, cartilage, and ciliated endoderm (black arrow) with secretory goblet cells (yellow arrow). **c**, Differentiated outgrowths from EpiSC embryoid bodies. Shown are markers indicative of neurons (neuronal specific β -III tubulin, Tubb3), astrocytes (glial fibrillary acidic protein, GFAP), cardiac muscle (cardiac specific troponin T, cTnT; and Nkx2.5), and skeletal muscle (Nebulin).

Recent studies have suggested that Oct3/4, Nanog and Sox2, in concert with interacting proteins, constitute an autoregulatory, pluripotent network^{11,19}. Dax1 (Nr0b1) has been identified as a central component of the protein interaction network and is required for the maintenance of mouse ES cells in culture^{19,20}. Although Oct3/4, Nanog and Sox2 were expressed at similar levels in mouse ES cells and EpiSCs, Dax1 was strikingly underexpressed in EpiSCs and human ES cells (Fig. 3a). Because EpiSCs and human ES cells differ significantly from mouse ES cells in their expression of this key component of the pluripotency network, we asked if the Oct3/4 targets differed. Genome-scale location analysis was used to identify Oct3/4 targets in mouse ES cells, EpiSCs and human ES cells. The data confirmed previous studies showing limited overlap of Oct3/4 targets between mouse ES cells and human ES cells^{11,12}, while showing a ~7-fold greater overlap of Oct3/4 targets between human ES cells and EpiSCs (Fig. 3c). These data emphasize that distinct transcriptional networks operate to maintain pluripotency in epiblast-like cells.

The origin of the germ line is an important issue in early embryology and in the biology of ES cells. Genes associated with the germ line, including *Stella*, *Piwil2*, *Stra8* and *Dazl* were expressed by mouse ES cells and significantly decreased or not detected in EpiSCs and human ES cells (Fig. 3a). This difference in the expression of germ cell specific genes is consistent with expression differences between the ICM and the post-implantation epiblast²¹. Primordial germ cell specification

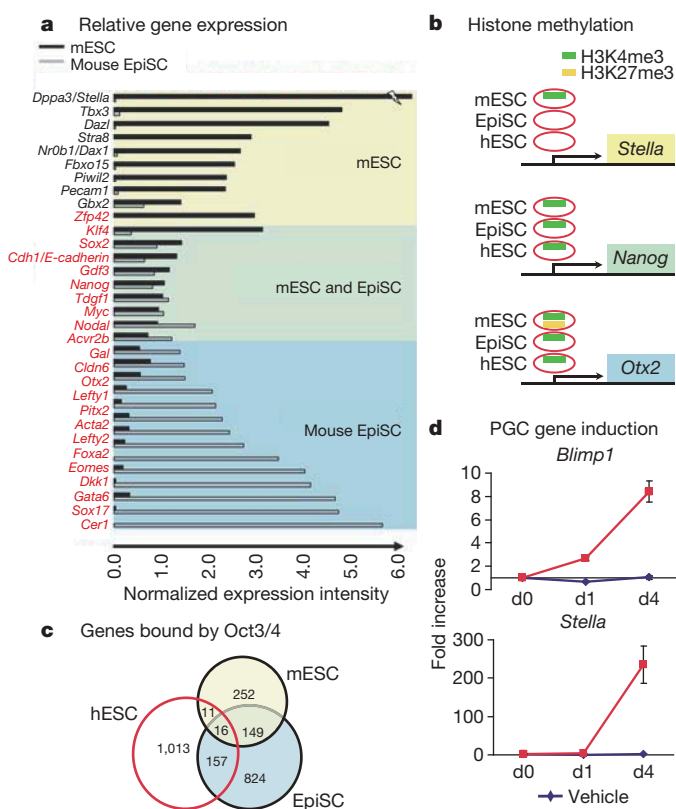


Figure 3 | Mouse ES cells and EpiSCs have distinct gene expression.

a, Normalized expression intensity values (scaled median ratio) were obtained from Agilent whole-genome microarrays. Gene names shown in red were detected in human ES cell cultures using Illumina and Agilent whole-genome microarrays. Three biological replicates were used for all three cell types. **b**, Representation of histone H3 methylation around the transcriptional start sites of *Stella*, *Nanog* and *Otx2* in mouse ES cells, EpiSCs and human ES cells (hESCs). The complete region can be viewed in Supplementary Fig. 7. **c**, Venn diagram of the number of genes bound by Oct3/4 in mouse ES cells, EpiSCs, and human ES cells using Agilent ChIP-on-chip promoter arrays. **d**, Real-Time PCR analysis of *Blimp1* and *Stella* expression following BMP4 treatment in EpiSCs. Data were obtained in triplicate and are presented as \pm s.d. PGC, primordial germ cell.

occurs in a precise temporal and spatial manner beginning with *Blimp1* (*Prdm1*) expression in a small number of cells in the posterior, proximal epiblast at E6.25 followed by upregulation of *Stella* around E7.25²¹. The growth factor BMP4 has a central role in primordial germ cell differentiation²². When EpiSCs were treated with BMP4, *Stella* and *Blimp1* were both induced in a temporal manner consistent with their specification *in vivo* (Fig. 3d). These data show that germ cell differentiation can be induced in EpiSCs and illustrates the value of EpiSCs as the proximal source of cells with restricted fates.

The expression of specific genes suggests that mouse ES cells and EpiSCs are two distinct pluripotent states representing cells of the pre-implantation embryo and later epiblast cells. Activin A (Inhb)/Nodal signalling is known to maintain the undifferentiated state of human ES cells as well as mouse post-implantation epiblast cells *in vivo*^{23–25}. In contrast, mouse ES cells require LIF signalling through the gp130/LIF receptor to maintain the pluripotent state²⁶. The undifferentiated state of human ES cells does not require LIF signalling²⁷. Like human ES cells, blocking STAT3 phosphorylation at tyrosine 705 with a JAK inhibitor supported the undifferentiated state of EpiSCs (Fig. 4a, b and Supplementary Fig. 8)²⁸. A selective inhibitor of type I activin

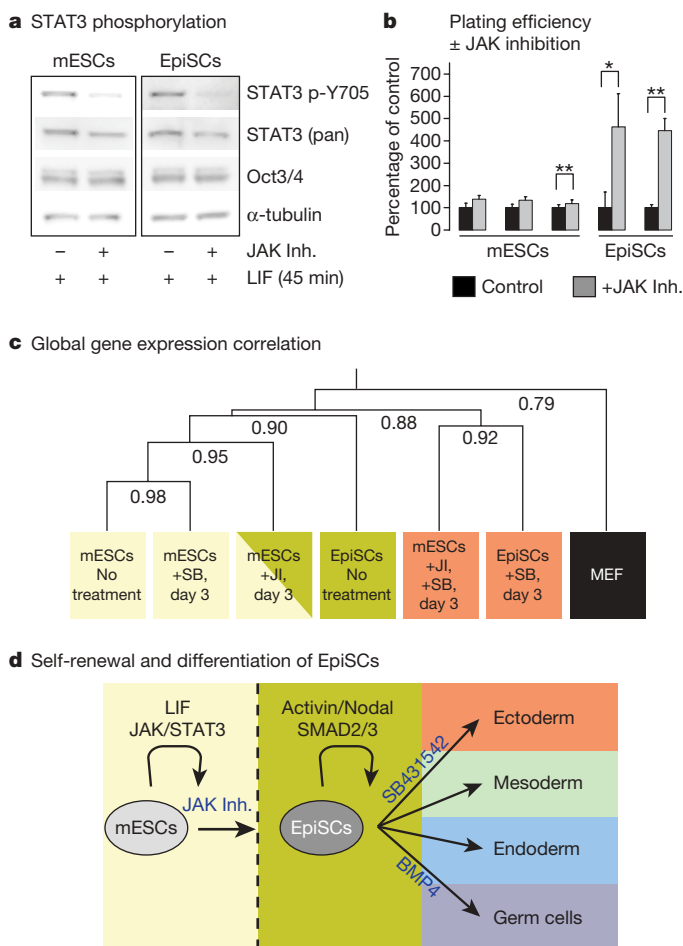


Figure 4 | Mouse ES cells and EpiSCs use distinct mechanisms to regulate pluripotency and differentiation.

a, Western blot showing JAK inhibitor abrogation of STAT3 phosphorylation at Y705 in response to LIF. **b**, Quantification of plating efficiency in response to JAK inhibition in three separate mouse ES cell and two separate EpiSC lines. Data are presented as a percentage of the vehicle-treated control \pm s.d. ($n = 4$). Statistically significant P values are noted with an asterisk (* $P < 0.05$ and ** $P < 0.01$). **c**, Whole-genome cluster analysis of transcripts from mouse ES cells and EpiSCs treated with the ALK inhibitor SB431542 (SB) and/or the JAK inhibitor (JI) for the indicated number of days. Correlation coefficients are indicated and represent the global level of similarity of gene expression among treated cell lines. **d**, Model of transitions in mouse ES cell and EpiSC developmental identity.

receptor-like kinases (ALK) 4, 5 and 7 (Acvr1b, Tgfb1 and Acvr1c, respectively) showed that the activin/Nodal pathway maintains pluripotency in EpiSCs. Activin/Nodal inhibition promoted differentiation of EpiSCs because colonies adopted a rosette-like morphology characteristic of neural precursors, were negative for pluripotency markers, and expressed genes specific to neuroectodermal fates in 3 days (*Sox1*, *Pax6* and *Nes*; Supplementary Table 1). An additional day of differentiation resulted in expression of genes found in more mature neural cells (doublecortin, *Dcx*; class III β -tubulin, *Tubb3*; and tyrosine hydroxylase, *Th*; Supplementary Table 1 and Supplementary Fig. 8). The differentiation of EpiSCs and mouse ES cells was analysed by whole-genome expression when pluripotency signals were blocked with small-molecule inhibitors. Mouse ES cells maintained their pluripotency markers even when the activin/Nodal pathway was inhibited (Fig. 4c, Supplementary Fig. 8, and Supplementary Table 1)²⁹. When both JAK and ALK inhibitors were used, mouse ES cells differentiated towards neuroectoderm and were most-closely related to EpiSCs treated with the ALK inhibitor alone (Figs 4c and Supplementary Table S1). Similar to EpiSCs, human ES cells also rapidly down regulated pluripotency genes and acquired neural characteristics when treated with the ALK inhibitor alone (Supplementary Fig. 8 and Supplementary Table 1). These data directly demonstrate that EpiSCs and human ES cells are functionally similar and specify their developmental relationship with mouse ES cells (Fig. 4c, d).

Although the use of human ES cells promises substantial advances in developmental biology and medicine, the differences between mouse and human ES cells have hindered their use. The pluripotent epiblast tissue is composed of the most immediate precursors of the early somatic lineages as well as the germ line. The important point established by this manuscript is that stable cell lines obtained from the post-implantation epiblast of the mouse embryo will aid in understanding the precise controls that regulate the transition from pluripotency to specific cell fates. There is increasing evidence that human ES cells differentiate to clinically relevant fates³⁰ and we propose a model that may accelerate efforts to realize their promise in medicine.

METHODS SUMMARY

Derivation and maintenance of cell lines from mouse post-implantation epiblast. Egg-cylinder-stage embryos in which the pro-amniotic cavity was restricted to the embryonic portion of the conceptus were obtained at E5.5 and dissected from their decidua in HEPES-buffered medium. Reichert's membrane was reflected with forceps or glass needles, and a cut was made with a glass scalpel at the boundary between the extra-embryonic ectoderm and the epiblast. The epiblast was isolated with a hand-pulled glass pipette after trypsin/pancreatin treatment. The isolated epiblast fragments were explanted one per 1.9 cm² well onto feeders in medium that consisted of DMEM-F12 (Invitrogen), 20% knockout serum replacement (KSR), 5 ng ml⁻¹ FGF2 (R&D Systems), 0.1 mM 2-mercaptoethanol (Sigma), 2 mM L-glutamine, and 1× non-essential amino acids (both from Invitrogen). EpiSCs were passaged every 2–3 days with collagenase type IV (Invitrogen) and triturated into small clumps.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 30 April; accepted 31 May 2007.

Published online 27 June 2007.

- Evans, M. J. & Kaufman, M. H. Establishment in culture of pluripotential cells from mouse embryos. *Nature* **292**, 154–156 (1981).
- Martin, G. R. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc. Natl Acad. Sci. USA* **78**, 7634–7638 (1981).
- Brook, F. A. & Gardner, R. L. The origin and efficient derivation of embryonic stem cells in the mouse. *Proc. Natl Acad. Sci. USA* **94**, 5709–5712 (1997).
- Chung, Y. *et al.* Embryonic and extraembryonic stem cell lines derived from single mouse blastomeres. *Nature* **439**, 216–219 (2006).
- Tesar, P. J. Derivation of germ-line-competent embryonic stem cell lines from preblastocyst mouse embryos. *Proc. Natl Acad. Sci. USA* **102**, 8239–8244 (2005).
- Gardner, R. L. Origin and differentiation of extraembryonic tissues in the mouse. *Int. Rev. Exp. Pathol.* **24**, 63–133 (1983).

- Surani, M. A., Hayashi, K. & Hajkova, P. Genetic and epigenetic regulators of pluripotency. *Cell* **128**, 747–762 (2007).
- Pelton, T. A., Sharma, S., Schulz, T. C., Rathjen, J. & Rathjen, P. D. Transient pluripotent cell populations during primitive ectoderm formation: correlation of *in vivo* and *in vitro* pluripotent cell development. *J. Cell Sci.* **115**, 329–339 (2002).
- Azuara, V. *et al.* Chromatin signatures of pluripotent cell lines. *Nature Cell Biol.* **8**, 532–538 (2006).
- Bernstein, B. E. *et al.* A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**, 315–326 (2006).
- Boyer, L. A. *et al.* Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122**, 947–956 (2005).
- Loh, Y. H. *et al.* The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nature Genet.* **38**, 431–440 (2006).
- Yeom, Y. I. *et al.* Germline regulatory element of Oct-4 specific for the totipotent cycle of embryonic cells. *Development* **122**, 881–894 (1996).
- Lawson, K. A., Meneses, J. J. & Pedersen, R. A. Clonal analysis of epiblast fate during germ layer formation in the mouse embryo. *Development* **113**, 891–911 (1991).
- Gardner, R. L., Lyon, M. F., Evans, E. P. & Burtenshaw, M. D. Clonal analysis of X-chromosome inactivation and the origin of the germ line in the mouse embryo. *J. Embryol. Exp. Morphol.* **88**, 349–363 (1985).
- Kubo, A. *et al.* Development of definitive endoderm from embryonic stem cells in culture. *Development* **131**, 1651–1662 (2004).
- Cai, J. *et al.* Assessing self-renewal and differentiation in human embryonic stem cell lines. *Stem Cells* **24**, 516–530 (2006).
- Liu, Y. *et al.* Genome wide profiling of human embryonic stem cells (hESCs), their derivatives and embryonic carcinoma cells to develop base profiles of U.S. Federal government approved hESC lines. *BMC Dev. Biol.* **6**, 20 (2006).
- Wang, J. *et al.* A protein interaction network for pluripotency of embryonic stem cells. *Nature* **444**, 364–368 (2006).
- Yu, R. N., Ito, M., Saunders, T. L., Camper, S. A. & Jameson, J. L. Role of Ahc in gonadal development and gametogenesis. *Nature Genet.* **20**, 353–357 (1998).
- Hayashi, K., de Sousa Lopes, S. M. & Surani, M. A. Germ cell specification in mice. *Science* **316**, 394–396 (2007).
- Lawson, K. A. *et al.* Bmp4 is required for the generation of primordial germ cells in the mouse embryo. *Genes Dev.* **13**, 424–436 (1999).
- Camus, A., Perea-Gomez, A., Moreau, A. & Collignon, J. Absence of Nodal signaling promotes precocious neural differentiation in the mouse embryo. *Dev. Biol.* **295**, 743–755 (2006).
- Vallier, L., Reynolds, D. & Pedersen, R. A. Nodal inhibits differentiation of human embryonic stem cells along the neuroectodermal default pathway. *Dev. Biol.* **275**, 403–421 (2004).
- James, D., Levine, A. J., Besser, D. & Hemmati-Brivanlou, A. TGF β /activin/nodal signaling is necessary for the maintenance of pluripotency in human embryonic stem cells. *Development* **132**, 1273–1282 (2005).
- Niwa, H., Burdon, T., Chambers, I. & Smith, A. Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3. *Genes Dev.* **12**, 2048–2060 (1998).
- Daheron, L. *et al.* LIF/STAT3 signaling fails to maintain self-renewal of human embryonic stem cells. *Stem Cells* **22**, 770–778 (2004).
- Androulakis-Theotokis, A. *et al.* Notch signalling regulates stem cell numbers *in vitro* and *in vivo*. *Nature* **442**, 823–826 (2006).
- Ogawa, K. *et al.* Activin–Nodal signaling is involved in propagation of mouse embryonic stem cells. *J. Cell Sci.* **120**, 55–65 (2007).
- Keller, G. Embryonic stem cell differentiation: emergence of a new era in biology and medicine. *Genes Dev.* **19**, 1129–1155 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature. A figure summarising the main result of this paper is included as Supplementary Figure 1.

Acknowledgements We thank J. Battey and Y. Ninomiya for comments on the manuscript and C. Graham, P. Fairchild, J. Pickel, K. Downs, W. Ferguson, and members of the McKay and Gardner labs and the NIH Stem Cell Unit for their contributions to this work. This research was supported by the Intramural Research Programs of the NIH–NINDS and NIH–NCI and by the Wellcome Trust. R.L.G. acknowledges the Royal Society for support. P.J.T. is an NIH–Oxford Biomedical Research Scholar.

Author Contributions P.J.T. derived the EpiSC lines. J.G.C. and P.J.T. carried out *in vitro* experiments. F.A.B., R.L.G. and P.J.T. derived the mouse ES cells. P.J.T., F.A.B. and T.J.D. performed chimera experiments. P.J.T. and D.L.M. performed teratoma experiments. P.J.T. and E.P.E. performed karyotypic analysis. P.J.T., J.G.C., and R.D.G.M. analysed the data and wrote the paper.

Author Information Microarray data are available at the Gene Expression Omnibus website (www.ncbi.nlm.nih.gov/geo) under accession number GSE7902. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to R.D.G.M. (mckayr@ninds.nih.gov) and P.J.T. (paultesar@ninds.nih.gov).

METHODS

Nomenclature. For clarity in this report, we refer to mouse ES cell lines derived from embryos prior to implantation in the uterus as 'mouse ES cells' and those reported herein from the epiblast of post-implantation mouse embryos as 'EpiSCs.'

Embryos. Natural matings were used for all experiments. Twelve hours from the middle of the dark period was termed embryonic day 0.5 (E0.5). Egg cylinder stage embryos in which the pro-amniotic cavity was restricted to the embryonic portion of the conceptus were obtained on E5.5 (± 4 h) from the 129SvEv (Taconic) or 129S2/SvHsd (Charles River) strains. Murine embryonic fibroblasts (MEFs) were obtained from E13.5 fetuses from the PO or CF1 strain. Morulae and blastocysts for chimera experiments were obtained from the MF1 or ICR strain on E2.5 or E3.5. PO or ICR females made pseudopregnant by mating with vasectomised males were used as recipients for embryo transfer on E2.5.

Mouse and human ES cell lines. Mouse ES cells mm10d, mm11e, mb1, ESF58/2, ESF122 and ESF175/1 were cultured on irradiated MEFs (feeders) with mouse ES cell medium, which consisted of Knockout DMEM (Invitrogen) supplemented with 15% fetal bovine serum (FBS) (Gibco) or Knockout Serum Replacement (KSR) (Invitrogen), 1×10^3 units ml^{-1} recombinant murine leukemia inhibitory factor (LIF) (ESGRO, Chemicon International), 2 mM L-glutamine, $1 \times$ non-essential amino acids (both from Invitrogen), and 0.1 mM 2-mercaptoethanol (Sigma). Mouse ES cells were passaged every third day as a single cell suspension using 0.25% trypsin/EDTA and seeded at 3.0×10^4 cells per cm^2 for routine culture. Human ES cell lines H9 (WiCell Research Institute; NIH Code WA09), hESBGN-03 (BresaGen, Inc.; NIH Code BG03), and HES-1 (ES Cell International; NIH Code ES01) were cultured on feeders with human ES cell medium, which consisted of DMEM-F12 (Invitrogen), 20% KSR, 5 ng ml^{-1} FGF2 (R&D Systems), 2 mM L-glutamine, 0.1 mM 2-mercaptoethanol, and $1 \times$ non-essential amino acids. Human ES cells were passaged every 4–6 days with enzymes (1 mg ml^{-1} collagenase type IV and 1 mg ml^{-1} dispase; both from Invitrogen) and triturated into small clumps of ~ 10 –100 cells. MEFs were irradiated with 30–60 gray and were seeded as feeders at a density of 4.8×10^4 cells per cm^2 and 1.88×10^4 cells per cm^2 for mouse and human ES cell cultures respectively. MEFs and feeders were maintained with DMEM supplemented with 10% FBS, 2 mM L-glutamine, and 0.1 mM 2-mercaptoethanol. All cells were grown on Nunclon Δ -treated dishes or multiwell plates (Fisher Scientific) coated for 2 h at 37 °C with 0.1% (w/v) gelatin (Sigma).

Derivation and maintenance of cell lines from mouse post-implantation epiblast. E5.5 embryos were dissected from their decidua in HEPES-buffered medium³¹. Reichert's membrane was reflected with forceps or glass needles and a cut was made at the boundary between the extra-embryonic ectoderm and the epiblast with a glass scalpel. The embryonic fragment containing the epiblast and overlying visceral endoderm was incubated for 8 mins at 4 °C in a solution of 0.5% trypsin and 2.5% pancreatin in PBS and subsequently returned to dissection medium. The embryos were drawn into a hand-pulled glass pipette with a diameter slightly smaller than the fragment to peel away the visceral endoderm³². The isolated epiblast fragments were rinsed in ES cell medium and explanted one per 1.9 cm^2 well. A variety of conditions were tested to maintain the epiblast explants in an undifferentiated, pluripotent state. Both mouse and human ES cell conditions were used along with the growth factors FGF2, activin A, BMP2 and BMP4 (all from R&D Systems), and murine LIF (Chemicon). The density of feeder cells was crucial to maintaining the epiblast cells in an undifferentiated state. A low feeder density resulted in differentiation into β -III tubulin⁺ neurons, whereas a high feeder density resulted in endoderm-like differentiation around the edges of the colonies. Addition of 10 ng ml^{-1} of activin A to epiblast explants on a low feeder density was sufficient to prevent neural differentiation. Higher concentrations of activin A resulted in differentiation. Addition of exogenous LIF did not drastically alter the cultures. Addition of 25 ng ml^{-1} BMP2 or BMP4 with or without LIF promoted differentiation. EpiSCs were optimally maintained with human ES cell passaging methods and medium containing FGF2 on a feeder density of 4.8×10^4 cells per cm^2 . EpiSCs were passaged every 2–3 days.

Small-molecule treatment. SB431542 was maintained as a 20 mM stock solution in DMSO (vehicle) and was provided to the cultures at the time of plating and every 24 h thereafter with the media change. JAK inhibitor I (Calbiochem 420099) was maintained as a 10 mM stock in DMSO and provided every 24 h.

Plating efficiency assays and subcloning. To obtain a pure, single cell suspension of mouse ES cells or EpiSCs, colonies were treated with 1.5 mg ml^{-1} collagenase for 10 mins and dislodged from the feeder layer with gentle pipetting. Colonies were settled away from any remaining feeders by gravity in a conical tube. Colonies were then trypsinized for 3 mins and counted. For plating efficiency assays, cells were plated onto feeders in 1.9 cm^2 wells at a density of 100 cells per cm^2 with vehicle or 0.6 μM JAK inhibitor I ($n = 4$ wells per

condition for each cell line examined). mm11e, ESF122 and ESF175/1 were used for the mouse ES cell samples and EpiSC-5 and EpiSC-7 were used for mouse EpiSC samples. Colonies were counted 4 days later. Results are presented as mean \pm standard deviation. Treatments that significantly differed from controls were identified using a paired, two-tailed Student's *t*-test (Microsoft Excel).

For subcloning, EpiSC-5 P22 cells were individually plated (1 cell per well) onto feeders in wells of a 96-well plate with the addition of 0.4 μM JAK inhibitor I. Five days later individual colonies were picked and dissociated in 8 μl drops of trypsin and plated into a fresh well of feeders. Clones were subsequently propagated using standard EpiSC passaging conditions.

Analysis of genomic integrity. Chromosome preparation and G-banding was carried out as described previously^{5,33}. For Comparative Genomic Hybridization (CGH) genomic DNA was isolated from cell lines or mouse liver using the Wizard Genomic DNA Purification Kit (Promega). Samples were processed for CGH analysis according to the Agilent Oligonucleotide Array-Based CGH Protocol Version 4.0. Briefly, 2.1 μg of genomic DNA was digested with *AluI* and *RsaI*. Digested DNA was labelled with either Cy3-dUTP or Cy5-dUTP using the Agilent Genomic DNA labelling Kit PLUS for 2 h, and purified with Microcon YM-30 filter units. Mouse 244A CGH arrays (Agilent) were hybridized with a labelled experimental sample and a labelled normal control sample. The control sample was isolated from liver of the same mouse strain as the experimental sample but was of opposite sex. The sex chromosome difference between the control and experimental samples was an internal control for the loss or gain of genetic material. Samples were hybridized for 40 h at 65 °C and washed with Agilent Oligo aCGH Wash Buffers 1 and 2. Processed arrays were scanned using an Agilent DNA microarray scanner controlled by Scanner Control Software v7.0 and the data were extracted using Agilent Feature Extraction Software v9.1. Data were analysed using CGH Analytics v3.4 (Agilent).

Mouse and human ES cells are capable of maintaining a normal karyotype in culture but are susceptible to some specific chromosomal abnormalities of the autosomes and the sex chromosomes. Human ES cells with an XX chromosome constitution are capable of maintaining both X-chromosomes even after extensive passaging, whereas XX mouse ES cells are unstable and typically lose one X during early passage³⁴. Independent EpiSC lines, EpiSC-3 (40, XX), EpiSC-5 (40, XY), and EpiSC-7 (40, XX) were karyotypically normal at passages 20, 17, and 22, respectively. At high passage (p33) an abnormal marker chromosome (a presumptive reciprocal translocation involving chromosomes 1 and 3) was observed in one preparation of line EpiSC-7. Owing to this, EpiSC-7 cells were not used beyond passage 26 for this study. The genetic integrity of the EpiSC-5 line and two derivative subclones was further examined by high-resolution comparative genomic hybridization (CGH). The parental line and one subclone (clone O) seemed normal, whereas the other subclone (clone L) showed an amplification of part of chromosome 8 (Supplementary Fig. 9); a common aberration previously found in mouse ES cells.

Differentiation. The potential of EpiSCs was tested *in vivo* by introducing the cells into blastocyst-stage embryos (lines EpiSC-2 and EpiSC-3). Methods for blastocyst injection were used as described previously^{5,35}. Chimaerism was judged by coat and eye color contribution. In addition, EpiSCs were transiently transfected with pMAX-GFP (Amara) and aggregated as small clumps with morula stage embryos in depression wells. The aggregates were cultured in KSOM and the incorporation/contribution of GFP⁺ EpiSCs was judged 24–36 hours later.

The developmental potential of EpiSC lines was assessed *in vitro* by forming embryoid bodies. Collagenase-isolated colonies were triturated into small clumps and grown for 1 wk in suspension using ultra-low attachment dishes (Corning). Embryoid bodies were subsequently plated 1 per well into gelatin- or fibronectin-coated 24-well plates. Culture medium consisted of DMEM with either 10% FBS or 15% KSR. For analysis of germ cell differentiation, EpiSC embryoid bodies were treated with 50 ng ml^{-1} BMP4 (R&D Systems) or vehicle in medium containing 10% FBS.

EpiSC lines were also tested for their ability to form teratomas in immunocompromised hosts. The inguinal mammary fat pads of female mice (athymic nu/nu, 23 days of age; NCI, Frederick, Maryland) were cleared of epithelium as described previously³⁶. EpiSC colonies were detached with collagenase and settled from feeders by gravity. Colonies were triturated to produce small cell clusters and these were resuspended in DMEM containing 10% FBS to 0.02×10^6 cells μl^{-1} . Cells (lines EpiSC-5 and EpiSC-7) were injected bilaterally into the cleared mammary fat pads (3 mice per line; 0.2×10^6 cells in 10 μl per fat pad). Tumours were allowed to develop for 3–6 weeks, at which time they were removed and fixed in 4% paraformaldehyde. Teratomas were embedded in paraffin, sectioned, and processed for immunohistochemistry or stained with haematoxylin and eosin. For immunohistochemistry, samples were deparaffinized with xylene, rehydrated and subjected to antigen unmasking (boiling in 0.01 M citrate for 5 mins). Slides were cooled in PBS and subsequently blocked in

10% serum. Primary antibody was added for 1.5 h at room temperature, the slides were rinsed, and secondary antibody was added for 45 mins. DAPI (Sigma, $1\mu\text{g ml}^{-1}$) was used to visualize nuclei. Slides were mounted with Shandon Immu-Mount (Thermo Scientific). Antibodies used were: mouse monoclonal IgG2a smooth muscle actin (clone 1A4; Sigma, 1:500), mouse monoclonal IgG1 nestin (clone Rat401; DSHB, 1:100), and mouse monoclonal IgG2a β -III tubulin (clone TUJ1; Covance, 1:500).

Immunostaining. Cells were prepared for immunostaining by fixation in 4% paraformaldehyde (Electron Microscopy Sciences) for 20 mins and subsequent permeabilization for 15 mins with 0.2% Triton-X in PBS. Cells were then blocked for non-specific binding with 10% normal goat or donkey serum (Abcam) in PBS for 2 h at room temperature. Primary antibodies were diluted in blocking solution and incubated with the samples overnight at 4 °C. Samples were rinsed with PBS and incubated with the appropriate fluorescently labelled secondary antibodies (Molecular Probes or Jackson ImmunoResearch, 1:500) for 2 h at room temperature. Primary antibodies used were: mouse monoclonal IgG2b Oct3/4 (C-10; Santa Cruz, 1:400), rabbit polyclonal Oct3/4 (H-134; Santa Cruz, 1:400), mouse monoclonal IgG1 Cdx2 (CDX2-88; Biogenex, 1:50), rabbit polyclonal Nanog (kindly provided by Dr. Ian Chambers, Edinburgh, U.K., 0.625 $\mu\text{g/ml}$), goat polyclonal Gata6 (AF1700; R&D Systems, 1.33 $\mu\text{g/ml}$), goat polyclonal Sox2 (AF2018; R&D Systems, 1.33 $\mu\text{g ml}^{-1}$), rabbit polyclonal nestin (McKay laboratory, 1:100), goat polyclonal Nkx2-5 (N-19; Santa Cruz, 1:50), mouse monoclonal IgG1 Troponin T (clone 13-11; Lab Vision, 1:100), mouse monoclonal IgG2a β -III tubulin (clone TUJ1; Covance, 1:500), rabbit polyclonal GFAP (Z0334; Dako, 1:400), goat polyclonal Dcx (C-18; Santa Cruz, 1:200), mouse monoclonal IgG1 nebulin (clone NB2; Sigma, 1:400), rabbit polyclonal von Willebrand factor (AB7356; Chemicon, 1:100), rabbit polyclonal TH (Pelfreeze, 1:200), mouse monoclonal IgG3 Sox17 (MAB1924, R&D Systems, 1.33 $\mu\text{g ml}^{-1}$), and rabbit polyclonal Nanog (Abcam, 1:400). Nuclei were visualized with DAPI (1 $\mu\text{g ml}^{-1}$).

Western blotting. Cultures were pre-incubated with JAK inhibitor I (0.6 μM) or vehicle and whole colonies were isolated, removed from feeders, and allowed to attach onto dishes coated with laminin (Roche) in DMEM supplemented with 15% KSR. After attachment, cultures were challenged with LIF (2×10^3 units ml^{-1}) for 45 mins and lysed in 2 \times SDS Lysis Buffer (1 \times , 50 mM Tris-HCl, pH 6.8, 100 mM DTT, 2% SDS, 0.1% bromophenol blue, and 10% glycerol). Equivalent amounts of protein extract were separated by SDS-PAGE (7% Tris-Acetate gels, Invitrogen) and transferred to a nitrocellulose membrane (0.45 μm pore, Invitrogen). Membranes were blocked with 5% milk in TBS and incubated with primary antibody overnight at 4 °C. Membranes were rinsed and incubated with HRP-conjugated secondary antibodies for 1 h at room temperature, followed by chemiluminescent detection using SuperSignal West Pico Substrate (Pierce). Primary antibodies used were: rabbit polyclonal STAT3 (C-20; Santa Cruz, 1:500), rabbit polyclonal STAT3 p-Y705 (9131; Cell Signalling, 1:1000), mouse monoclonal Oct3/4 (C-10; Santa Cruz, 1:1000), and mouse monoclonal α -tubulin (clone B-5-1-2; Sigma, 1:6000).

RNA preparation and RT-PCR. Cultures were treated with 1.5 mg ml^{-1} collagenase in KSR-containing medium for 10 mins at 37 °C. The collagenase solution was removed and fresh medium was added to the cultures. Colonies were dislodged from the feeder layer, if present, by gentle pipetting. Colonies were settled away from any remaining feeder cells by gravity in a conical tube. Total RNA was prepared using TRIzol (Invitrogen) and samples were treated with DNase (DNA-free, Ambion). RNA (200 ng) was reverse-transcribed using Superscript III (Invitrogen) with random primers. For semi-quantitative RT-PCR, complementary DNA was amplified for 30 cycles with Taq polymerase (Roche) using gene specific primers, and products were analysed on ethidium-bromide-stained agarose gels. All primer sequences except those for actin were adopted from ones described previously³⁷: *Oct4*, forward 5'-CGTTCTCTT-TGGAAAGGTGTTTC-3', reverse 5'-GAACCATACTCGAACCACATCC-3'; *Zfp42(Rex1)*, forward 5'-TGAAAGTGAGATTAGCCCGAG-3', reverse 5'-GTCCCATCCCCTTCAATAGCAC-3'; *Fgf5*, forward 5'-CTGTACTGCAGAG-TGGGCATCGG-3', reverse 5'-GACTTCTGCGAGGCTGCGACAGG-3'; *actin*, forward 5'-CTAGACTTCGAGCAGGAGATGGC-3', reverse 5'-TCTGCATC-CTGTGACGAATGCC-3'. Real-time PCR was performed according to the manufacturer's protocol using the Applied Biosystems (ABI) 7900HT instrument with the ABI TaqMan Universal PCR Master Mix and Taqman pre-optimized gene expression assays (*Prdm1(Blimp1)*, Mm00476128_m1; *Dppa3 (Stella)*, Mm00836373_g1; *Gapdh*, 4352662).

Reporter gene assays. For the *Oct3/4 DE-SV40-Luc* and *Oct3/4 PE-SV40-Luc* constructs, mouse genomic DNA containing the previously described *Oct3/4* enhancers¹³ was amplified and inserted into the pCR2.1-TOPO vector (Invitrogen). The following primers were used to amplify the appropriate genomic fragments: DE, forward 5'-ACGTAAGTACTTCAGACACCAGAAG-AGG-3', reverse 5'-GAGCCTACCCTGAACCTCCCTAGAGTG-3'; PE, forward

5'-CACTCTAGGGAAGTTCAGGGTAGGCTC-3', reverse 5'-AAAGCCTG-TTGGCACTGCACCCTCTCGG-3'. The inserts were cut from the pCR2.1-TOPO vector using *XhoI* and *KpnI* and inserted into the pGL3-Promoter Vector (Promega), which was digested with *XhoI* and *KpnI* using standard cloning techniques. All constructs were sequenced. For the luciferase assays, equimolar amounts of each construct were transfected into 2.5×10^6 cells using the Amaxa Nucleofection kit along with the pRL-TK vector for normalization. Assays were performed 24–48 hours later using the Dual-Glo Luciferase Assay System (Promega) and signals were detected using a Wallac Victor3 Plate Reader (Perkin Elmer). The basal activity of the empty luciferase vector was set as 1.0.

Expression microarrays. Agilent: Cy3-CTP labelled cRNA was produced from 200 ng of total RNA using the Agilent Low RNA Input Liner Amplification Kit. After labelling and cRNA purification, cRNA was quantified using the NanoDrop ND-1000 UV-VIS Spectrophotometer. Cy3- labelled cRNA (1.65 μg) was hybridized per individual 44K array on catalogue Agilent 4x44K Mouse (G4122F) or Human (G4112F) whole-genome microarrays for 17 h at 65 °C (10 r.p.m.). After hybridization, arrays were washed consecutively with Agilent gene expression wash buffers one and two, and acetonitrile for 1 min each, followed by a final 30 second wash in Agilent Stabilization and Drying solution. Slides were scanned immediately following washing in the Agilent Scanner using Scan Control 7.0 software. The scan resolution was 5 μm and the eXtended Dynamic range feature (XDR Hi 100%, XDR Low 10%) was used to avoid saturated features. Scan data were extracted with Feature Extraction 9.1 software using the GE1-V5_91 protocol. Extracted signal intensities were analysed using GeneSpring 7.3.1 software and the datasets were normalized using standard Agilent FE Import 1-color settings. This normalization procedure accounted for interchip variability by adjusting the intensity distribution to the 50th percentile.

Comparison of mouse ES cells and EpiSCs was done using 3 biological replicates for each cell type. For mouse ES cells, lines ESF175/1 p16, ESF58/2 p15, and ESF122 p12 were used and EpiSC-7 P20, EpiSC-5 P22, and EpiSC-7 p25 were used for the EpiSC samples. The expression of each gene in the mouse ES cell dataset was compared to the matching gene in the EpiSC dataset through a per gene median scaling. Resulting normalized intensity values equal to 1.0 indicated that the gene was expressed equally in the two cell types, whereas values greater than 1.0 indicated elevated gene expression and values less than 1.0 indicated decreased expression as compared to the other cell type. Human ES cell lines H9 P52, ES01 P66, and BG03 P49 were used as the 3 biological replicates for the human ES cell dataset. Because no comparison was made with other human cell types, the per gene median scaling was excluded in the human ES cell analysis.

When comparing the effect of SB431542 treatment on mouse ES cells and EpiSCs two separate biological replicates were used for the analysis (ESF122 and ESF175/1 for mouse ES cell samples, and EpiSC-5 and EpiSC-7 for mouse EpiSC samples). For each cell type, time-course samples were normalized to the non-treated samples and therefore all non-treated samples had a value of approximately 1.0. The normalized values for genes up regulated by the SB431542 were greater than 1.0 and those decreased by the treatment were less than 1.0. Gene lists of statistically significant ($P < 0.05$) 2.5-fold up regulated or 2.5-fold down regulated genes were obtained from GeneSpring. Analysis of SB431542 treatment of human ES cells was performed as above for cell line H9 P43.

Average linkage cluster maps and correlation coefficients between transcript expression datasets were generated using Genespring software. A maximal value of 1.0 indicates equivalent detection of over 41,000 genes and transcripts represented on the Agilent mouse whole-genome oligo microarray.

Illumina: the Illumina TotalPrep RNA Amplification Kit (Ambion) was used to amplify 100 ng of total RNA from each sample. cRNA was quantified using the Quant-iT RiboGreen RNA Reagent (Invitrogen), and 1.5 μg of material was hybridized to Illumina Sentrix Human-6 Expression Beadchips for 17 h at 55 °C. Beadchips were washed and stained according to the protocol described in the Whole-Genome Gene Expression for Beadstation Manual Revision D. The Beadchips were scanned in an Illumina BeadStation controlled by BeadScan software. Data were analysed using the gene expression module of BeadStudio 2.3.33 software and datasets were normalized using rank invariant normalization. Illumina arrays provide bead redundancy and therefore each gene or transcript is represented an average of 30 times on each array. Owing to this, a detection P value can be calculated for each gene or transcript. Detection P values less than 0.01 were considered significant. Human Sox2 expression was only analysed on the Agilent arrays because there is a known defect with the Sox2 bead on this version of the Illumina bead array.

Chromatin immunoprecipitation on DNA microarrays (ChIP-on-chip). ChIP-on-chip was performed using the Agilent mammalian protocol with modification. To immunoprecipitate chromatin, 5×10^7 cells were treated with 1% formaldehyde at room temperature for 10 mins followed by exposure to 0.125 M glycine and a subsequent wash in PBS. Cells were lysed in Cell Lysis Buffer (5 mM

HEPES, pH 8.0, 85 mM KCL, 0.5% Triton) for 5 mins on ice and the nuclei were re-suspended in Nuclei Lysis Buffer (50 mM Tris-HCL, pH 8.0, 10 mM EDTA, 1% SDS). Lysates were sonicated under conditions yielding fragments ranging from 200 bp to 1,000 bp. Samples were subsequently pre-cleared at 4 °C with recombinant protein G agarose beads (Invitrogen) coated in yeast transfer RNA for 1 h. Three per cent of the sonicated material was reserved as whole-cell extract material. Pre-cleared lysate was diluted 1:10 in IP buffer (0.01% SDS, 1.1% Triton, 1.2 mM EDTA, 16.7 mM Tris-HCL, pH 8.0, 167 mM NaCl) for an overnight immunoprecipitation with specific antibody at 4 °C with rocking. Antibodies used were: Oct3/4 (Santa Cruz sc-9081x Lot no. J1703), trimethyl-histone H3 K4 (Abcam no. ab8580 Lot no. 122062) or trimethyl-histone H3 K27 (Abcam no. ab6002 Lot no. 222266). We used 1.5 µg of antibody per 1×10^6 cells. Complexes were collected for 4 hours using recombinant protein G agarose beads (Invitrogen) coated in BSA and yeast tRNA. Samples were washed for 10 mins in each of the following buffers; IP dilution buffer, Dialysis Buffer (2 mM EDTA, 50 mM Tris-HCL, pH 8, 0.2% Sarkosyl), TSE-500 (0.1% SDS, 1% Triton, 2 mM EDTA, 20 mM Tris-HCL, pH 8, 500 mM NaCl), LiCl Wash (100 mM Tris-HCL, pH 8, 500 mM LiCl, 1% Triton, 1% Deoxycholic acid), and TE Buffer. After washing, samples were eluted in Elution Buffer (50 mM NaHCO₃, 1% SDS), and formaldehyde cross-links were reversed by an overnight incubation at 65 °C. Samples were purified through PCR Purification Kit columns (Qiagen). Immunoprecipitation-enriched and whole-cell extract DNA was blunted using the DNA Terminator End Repair Kit (Lucigen) before linker ligation and amplification, which was performed according to the Agilent protocol. Labelling was performed using the Bioprime DNA Labeling System (Invitrogen) and Cy3-dUTP or Cy5-dUTP (Amersham). Human G4489A and mouse G4490A 244K promoter arrays (Agilent) were probed for 40 h at 65 °C and subsequently scanned using an Agilent Scanner controlled by Agilent Scan Control 7.0 software. Custom 244K mouse and human ChIP-on-chip arrays were designed using eArray (Agilent) and processed as described above. Data were extracted with Agilent Feature Extraction 9.1 software. Oct3/4-bound genes were called using Agilent's ChIP Analytics 1.3 software, incorporating the Whitehead Error Model, which has a false-positive rate of approximately 0.5% and a false-negative rate of approximately 20% (ref. 11). A Venn diagram was created for overlapping and non-overlapping Oct3/4-bound genes in mouse ES cells, mouse EpiSCs and human ES cells. Genes were only included in the comparison if they had a homologue on the other species' array. For example, probe data for the *Nanog* gene were not present in the mouse array dataset and therefore *Nanog* was not included in the cross-species comparison. Probes that were associated with modified histones were called on the basis of a neighbourhood *P*value of less than 0.01 (Agilent ChIP Analytics 1.3 Software). Mapping of bound probes was performed in the UCSC genome browser using genome build mm7 (August 2005) for mouse and HG17 (May 2004) for human³⁸.

31. Downs, K. M. *In vitro* methods for studying vascularization of the murine allantois and allantoic union with the chorion. *Methods. Mol. Med.* **121**, 241–72 (2006).
32. Beddington, R. S. *Isolation, culture and manipulation of post-implantation mouse embryos* (ed. Monk, M.) (IRL Press, Oxford, 1987).
33. Nagy, A. *Manipulating the mouse embryo: a laboratory manual* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 2003).
34. Zvetkova, I. *et al.* Global hypomethylation of the genome in XX embryonic stem cells. *Nat. Genet.* **37**, 1274–1279 (2005).
35. Robertson, E. J. *Teratocarcinomas and embryonic stem cells: a practical approach* (IRL, Oxford; Washington DC, 1987).
36. Deome, K. B., Faulkin, L. J. Jr., Bern, H. A. & Blair, P. B. Development of mammary tumors from hyperplastic alveolar nodules transplanted into gland-free mammary fat pads of female C3H mice. *Cancer. Res.* **19**, 515–520 (1959).
37. Hirst, C. E. *et al.* Transcriptional profiling of mouse and human ESCs identifies *SLAIN1*, a novel stem cell gene. *Dev. Biol.* **293**, 90–103 (2006).
38. Kent, W. J. *et al.* The human genome browser at UCSC. *Genome. Res.* **12**, 996–1006 (2002).

LETTERS

An intracellular P2X receptor required for osmoregulation in *Dictyostelium discoideum*

Samuel J. Fountain¹, Katie Parkinson¹, Mark T. Young¹, Lishuang Cao¹, Christopher R. L. Thompson¹ & R. Alan North¹

P2X receptors are membrane ion channels gated by extracellular ATP^{1,2} that are found widely in vertebrates, but not previously in microbes. Here we identify a weakly related gene in the genome of the social amoeba *Dictyostelium discoideum*, and show, with the use of heterologous expression in human embryonic kidney cells, that it encodes a membrane ion channel activated by ATP (30–100 μ M). Site-directed mutagenesis revealed essential conservation of structure–function relations with P2X receptors of higher organisms. The receptor was insensitive to the usual P2X antagonists³ but was blocked by nanomolar concentrations of Cu²⁺ ions. In *D. discoideum*, the receptor was found on intracellular membranes, with prominent localization to an osmoregulatory organelle, the contractile vacuole. Targeted disruption of the gene in *D. discoideum* resulted in cells that were unable to regulate cell volume in hypotonic conditions. Cell swelling in these mutant cells was accompanied by a marked inhibition of contractile vacuole emptying. These findings demonstrate a new functional role for P2X receptors on intracellular organelles, in this case in osmoregulation.

The *D. discoideum* genome contains five sequences predicted to encode proteins homologous to vertebrate P2X receptors. The protein most closely related to the seven human receptors is DDB0168616 (Supplementary Fig. 1). HEK cells expressing a humanized version of this complementary DNA responded to ATP with robust inward currents (Fig. 1). ATP evoked unitary currents (conductance 8.2 pS at –100 mV) in outside-out patches (Fig. 1a, b), establishing that this gene (*D. discoideum* p2xA) encodes an ATP-gated ion channel (*DdP2X*). Whole-cell currents evoked by ATP were concentration-dependent (10–300 μ M) with kinetic properties most similar to those observed for human P2X₂ or P2X₄ receptors¹ (Fig. 1d). With sodium as the only cation, the current–voltage relation reversed at zero and showed moderate inward rectification (Fig. 1c). By measuring the current reversal potentials in different extracellular ions, we found that the channels were freely permeable to Ca²⁺ ions but less so to larger cations (the relative permeability P_X/P_{Na} was 1.5 ± 0.11 ($n = 8$ cells), 0.54 ± 0.01 (5), 0.48 ± 0.21 (5), 0.38 ± 0.16 (5) and 0.21 ± 0.05 (8), where X represents Ca²⁺, choline, Tris, tetraethylammonium and *N*-methyl-D-glucamine, respectively; Fig. 1c). These values are in the range of vertebrate P2X receptors¹. Western blotting of the expressed receptor showed two bands at 48 and 55 kDa (Fig. 1e). The receptor formed a trimer under non-denaturing conditions (Fig. 1e), and immunocytochemistry indicated that the expressed receptor was clearly localized at the plasma membrane (Fig. 1f). These data are consistent with a body of evidence indicating that vertebrate P2X receptors are trimeric membrane proteins^{1,2,4,5}.

The *DdP2X* receptor was strongly activated by hydrolysis-resistant ATP analogues. $\beta\gamma$ -Imido-ATP was about tenfold more potent than ATP, and $\alpha\beta$ -methylene-ATP was equipotent with ATP

(Fig. 2a, b). The receptor was only weakly activated by 2',3'-O-(4-benzoylbenzoyl)-ATP, and not at all by UTP, CTP, GTP, ADP, cyclic AMP, NAD⁺, FAD, ADP-ribose and cyclic ADP-ribose (1 mM). Most P2X receptors are blocked by suramin and pyridoxal-phosphate-6-azophenyl-2',4'-disulphonate, whereas P2X₁, P2X₃ and P2X_{2/3} heteromeric receptors are blocked by nanomolar concentrations of 2',3'-O-(2,4,6-trinitrophenyl)-ATP^{1,3}: none of these antagonists inhibited the currents at *DdP2X* receptors, even at 30–100 μ M (Fig. 2c). However, *DdP2X* was very sensitive to block by Cu²⁺ (half-maximal inhibitory concentration (IC₅₀) 40 ± 5 nM, $n = 5$; Fig. 2c), whereas other divalent ions were much less effective (Ni²⁺ IC₅₀ 59 ± 5 μ M; Ca²⁺ IC₅₀ >20 mM; $n = 5$). Inhibition by micromolar Cu²⁺ has been described for some (P2X₄, P2X₇) but not other mammalian P2X receptors^{6,7}.

We tested the essential similarity in molecular physiology between the *DdP2X* receptor and its vertebrate counterparts by mutating residues known to be critically involved in function^{1,8,9}. A body of work implicates two lysine residues in the receptor ectodomain as contributing to ATP binding. The first of these (*DdP2X* Lys 67) is in a position equivalent to Lys 71 of the rat P2X₂ receptor (Supplementary Fig. 1); mutation to alanine (*DdP2X*[K67A]) caused a large (more than tenfold) decrease in ATP sensitivity. The second, *DdP2X* Lys 289, aligns with rat P2X₂ receptor Lys 308; ATP elicited much smaller currents at *DdP2X*[K289A] (Fig. 2d, e). An aspartate residue in the second transmembrane domain is essential for function in vertebrate P2X receptors¹; the *Dictyostelium* receptor in which this was replaced by alanine [D330A] was also non-functional. Finally, a YXXXX motif at the carboxy terminus of all vertebrate P2X receptors is involved in receptor stabilization at the plasma membrane¹⁰; this motif was also required in *DdP2X* for robust membrane currents (Fig. 2d). Conversely, some regions of the receptor that have been deduced to be critical for ATP binding and/or channel activation in mammalian P2X receptors were clearly not essential. These included the conserved NFF Φ Φ KNS Φ (Φ being any hydrophobic residue) and GYNFRFAKY motifs in the ectodomain of mammalian receptors. The former motif is found in *DdP2X* (SFTILIDHTM), but in contrast with the human P2X₁ receptor⁹ the phenylalanine residue (Phe 156) was not required for ATP action (Fig. 2d, e). The latter motif is not present in the *DdP2X* sequence.

A green fluorescent protein (GFP)-tagged version of *DdP2X* (*DdP2X*–GFP) was localized to intracellular membranes of *D. discoideum* cells rather than the plasma membrane (Fig. 3a), contrary to what is observed either for P2X receptors in mammalian cells or for *DdP2X* receptors expressed in HEK 293 cells (Fig. 1f). Strong colocalization with calmodulin demonstrated prominent expression on the membrane of the contractile vacuole (Fig. 3a), a specialized osmoregulatory organelle that allows protists to survive in hypotonic environments such as soil and fresh water^{11–14}. Identification of *DdP2X* as a contractile vacuole protein thus indicated that *DdP2X* receptors might be involved in osmoregulation. To test this, we

¹Faculty of Life Sciences, University of Manchester, Michael Smith Building, Oxford Road, Manchester M13 9PT, UK.

generated mutant cells with a disrupted *p2xA* gene by homologous recombination (Fig. 3b) and assayed for osmoregulatory defects. In hypotonic conditions, wild-type cells swelled by about 20%, followed by a regulatory decrease in cell volume over the next 40 min (Fig. 3c, d; $n = 30$ cells). In contrast, *DdP2X*-null cells showed no regulatory decrease in cell volume but continued to swell over 60 min (Fig. 3; $n = 30$ cells, $P < 0.01$). Two lines of evidence indicate that this phenotype is due to a requirement for *DdP2X* receptors in wild-type cells. First, expression of *DdP2X*-GFP in the *DdP2X*-null cells rescued the osmoregulatory defects of the mutant (Fig. 3d). Second, Cu^{2+} at concentrations that blocked *DdP2X* currents when expressed in HEK 293 cells (Fig. 2) recapitulated the swelling phenotype in wild-type cells (Fig. 3d). Furthermore, $\beta\gamma$ -imido-ATP (100 μM), which is polar and cell-impermeant, had no effect on the regulatory decrease in volume when applied to wild-type cells; this discounts a

role for plasma membrane receptors and is consistent with the intracellular compartmentalization of *DdP2X*. Taken together, these data indicate a direct role for *DdP2X* receptors in osmoregulation.

In wild-type cells, the contractile vacuole periodically fills and voids, expelling water from the cell body and permitting the regulatory decrease in volume under hypotonic conditions. We therefore tested whether an impairment of contractile vacuole function could explain why *DdP2X*-disrupted cells displayed no regulatory decrease in volume. In wild-type cells, regular voiding events occurred with a cycle of 80 ± 5 s ($n = 30$ cells; Fig. 3e). In contrast, mutant cells had a decreased frequency of contractile vacuole voiding and a prolonged cycle (182 ± 10 s, $n = 30$ cells; Fig. 3e). Wild-type cells treated with 10 μM Cu^{2+} also had a prolonged voiding cycle (202 ± 13 s, $n = 15$

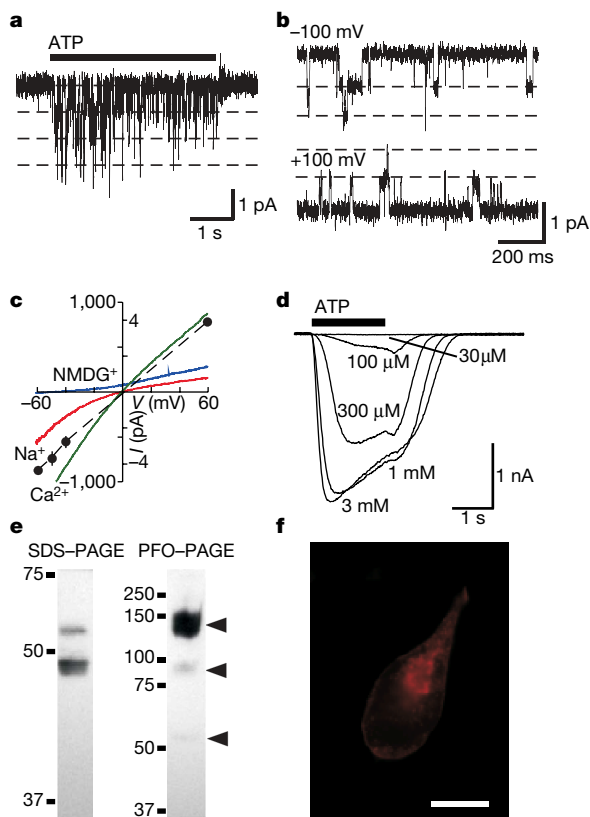


Figure 1 | *DdP2X* receptor is an ATP-gated ion channel. **a, b**, Single-channel openings in response to 100 μM ATP recorded from outside-out patches taken from transfected HEK 293 cells. **c**, Whole-cell current-voltage relationship shows modest inward rectification (red trace). Replacement of extracellular sodium by *N*-methyl-D-glucamine (each 147 mM) shifts reversal potential from 0 mV to -50 mV (blue trace). The current reverses close to zero with isotonic extracellular Ca^{2+} (green trace). Filled circles indicate amplitudes of unitary currents recorded in outside-out patches. Error bars indicate s.e.m. ($n = 6$). **d**, Concentration-dependent responses to ATP. Typical whole-cell recordings of inward currents evoked by different concentrations of ATP applied for 2 s, repeated at 2-min intervals. Cells were held at -60 mV. **e**, SDS-PAGE western blot analysis of whole-cell lysates from HEK 293 cells expressing His-tagged *DdP2X* shows bands at 48 and 55 kDa. The mass of both bands was decreased by treatment with PNGaseF (not shown), indicating that both were *N*-glycosylated, and both bands were present at the cell surface as measured by cell-surface biotinylation. His-tagged *DdP2X* migrates on perfluoro-octanoic acid (PFO)-PAGE predominantly as a trimer (although dimeric and monomeric species are also detectable, as indicated by the middle and lower arrows). Numbers at the left of the gels are molecular masses in kDa. **f**, Immunocytochemistry showing predominant localization of the His-tagged *DdP2X* receptor (red; secondary antibody was Cy3) at the membrane in HEK 293 cells. Scale bar, 10 μm .

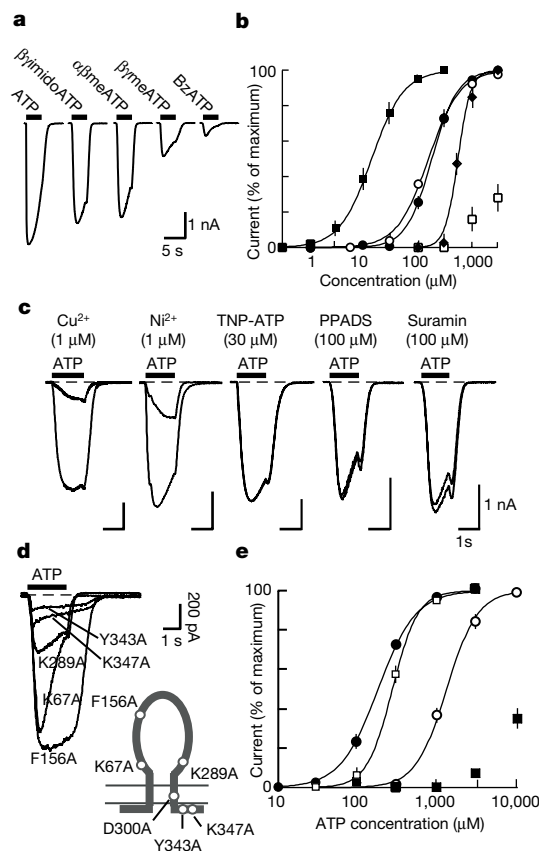


Figure 2 | Properties of *DdP2X* receptors. **a**, Inward currents evoked by ATP and analogues (each 300 μM , for 2 s) in HEK 293 cells expressing *DdP2X* receptors (holding potential -60 mV). **b**, Concentration-response curves for experiments such as those shown in **a**. Responses are normalized to each maximum response and show a rank order of potency $\beta\gamma$ -imido-ATP (filled squares) $>$ $\alpha\beta$ -meATP (open circles) = ATP (filled circles) $>$ $\beta\gamma$ -meATP (filled diamonds) $>$ benzoyl-benzyl-ATP (BzATP; open squares); each point is the mean \pm s.e.m. for five to ten cells. **c**, Responses to ATP (100 μM) before and after application of Cu^{2+} , Ni^{2+} and P2X antagonists. The duration of preapplication of antagonists was 2 min (Cu^{2+} and Ni^{2+}) or 4 min (2',3'-O-(2,4,6-trinitrophenyl)-adenosine 5'-triphosphate (TNP-ATP), pyridoxal-phosphate-6-azophenyl-2',4'-disulphonate (PPADS) and suramin). **d**, Representative whole-cell currents evoked by ATP (3 mM) in HEK 293 cells expressing mutant receptors. Lower right: diagram illustrating the topology of the P2X receptor and the relative positions of the residues mutated. **e**, Comparison of ATP sensitivity between wild-type *DdP2X* receptors (filled circles) and the K67A (open circles), K289A (filled squares) and F156A (open squares) mutations. Each point is the mean \pm s.e.m. for five to ten cells. The immunohistochemical appearance of cells expressing mutated receptors (including D330A) with the anti-His antibody was the same as that of wild-type cells. Biotinylation and SDS-PAGE followed by gel densitometry showed a highly significant expression of wild-type, K67A and D330A receptors at the cell surface.

cells). These results suggest that loss of the *DdP2X* receptor leads to a compromised contractile vacuole, an inability to osmoregulate, and ultimately cell swelling. Contractile vacuole filling seemed to be normal in *DdP2X*-null cells, implying that docking at the cell membrane or the contraction itself is deficient. It is possible that the Ca^{2+} permeability of the P2X receptor in the membrane of the contractile vacuole has a key function. In concord with this idea, the phenotype of *DdP2X*-disrupted cells is similar to that of *LvsA*-null cells, which is thought to be partly due to a misregulation of calmodulin function^{15,16}. One speculative possibility is that distension of the contractile vacuole leads to the entry of ATP from the cell cytoplasm: this then activates *DdP2X* receptors, which allow vacuolar Ca^{2+} ions to pass back into the cytoplasm and initiate contraction.

These studies represent significant advances for the study of P2X receptor structure and function. The *DdP2X* receptor reported here is relatively remote in sequence from those previously reported^{1,17}. Despite this, essential properties of a P2X receptor (ATP binding, channel gating and permeation) are conserved between amoeba and man. Consequently, attention can now be drawn to those few parts of

the molecule critical to these functions. P2X receptors have hitherto been considered in the context of the plasma membrane and the transmitter-like actions of extracellular ATP^{1,2}. The present studies, however, define a receptor that is both found on internal membranes and has an intracellular role. This raises the possibility that P2X receptors might also function as intracellular ion channels in other cell types, including higher organisms. Such an idea is not without precedent. Another family of channel proteins (ClC chloride channels) has members with well-defined functional roles in both plasma and endosomal membranes¹⁷. Furthermore, some members of the family of 'ectoenzyme' nucleoside triphosphate diphosphohydrolases are actually found predominantly on internal membranes¹⁸. Given the ubiquitous presence of their ligand ATP, P2X receptors may be more general regulators of intracellular organelles in ways other than osmoregulation.

METHODS

Generation and expression of *DdP2X* constructs. A 1.1-kilobase cDNA corresponding to the predicted open reading frame of DDB0168616 was amplified by polymerase chain reaction from growth-phase AX4 strain *D. discoideum*. The insert was subcloned into pcDNA3.1-His at 5' *Bam*HI and 3' *Xho*I sites and into pDEX 27 GFP at *Eco*RI sites for expression in HEK 293 cells and *D. discoideum*, respectively. Expression of the pcDNA3.1 *DdP2X* construct in HEK 293 cells produced messenger RNA transcript but no detectable protein or function. A humanized version of the *DdP2X* receptor was synthesized in which each amino acid codon was replaced by that most commonly found in *Homo sapiens*. This was C-terminally tagged with hexahistidine (DNA 2.0 Inc.) and subcloned into pcDNA3.1. For electrophysiology, HEK 293 cells were transiently transfected with 1 μg (whole-cell) or 0.1 μg (outside-out) plus 0.1 μg of enhanced GFP using Lipofectamine 2000. For expression in *D. discoideum*, cells were transformed by electroporation and subjected to G418 selection.

Electrophysiology. Whole-cell and outside-out patch recordings were made from HEK 293 cells at room temperature (18–20 °C), 24–48 h after transient transfection. The extracellular solution contained (in mM): 145 NaCl, 2 KCl, 2 CaCl_2 , 1 MgCl_2 , 13 D-glucose and 10 HEPES, pH 7.3. For whole-cell recordings, the intracellular patch pipette solution contained (in mM): 145 NaCl, 10 HEPES and 10 EGTA, pH 7.3. For outside-out recordings, NaF replaced NaCl in the patch pipette. Outside-out recordings were sampled at 10 kHz and filtered at 3 kHz. For monovalent ion substitution in whole-cell experiments $P_{\text{X}}/P_{\text{Na}}$ was calculated from $\exp(\Delta E_{\text{rev}}F/RT)$, and for Ca^{2+} $P_{\text{Ca}}/P_{\text{Na}}$ was from $[\text{Na}]_i \{ \exp(E_{\text{rev}}F/RT) / (1 + \exp(E_{\text{rev}}F/RT)) \} / 4 [\text{Ca}^{2+}]_o$, where $[\text{Na}]_i$ is the cytosolic Na_i concentration, F is Faraday's constant, R is the universal gas constant, T is the absolute temperature and $[\text{Ca}^{2+}]_o$ is the extracellular Ca^{2+} concentration. E_{rev} values were corrected for calculated liquid junction potentials. Numerical data are presented as means \pm s.e.m.

Osmoregulation assay and contractile vacuole voiding. Wild-type and mutant AX4 strain cells were maintained in HL5 growth medium at 22 °C. Cells adherent to glass coverslips were imaged with an inverted Olympus IX71 microscope (40 \times objective, 1.6 numerical aperture); images were captured at 2-min intervals with Simple PCI software (C-Imaging Systems). Geometric cell size was measured offline for ten cells per field over three independent experiments. Contractile vacuoles were observed at 1-s intervals for 15 min after exposure to water, and voiding time was measured directly.

Immunoblotting and immunostaining. We used 50 μg of total cell lysate per load for perfluoro-octanoic acid PAGE¹⁹ (8%) and SDS-PAGE (4–12% gradient; NuPAGE; Invitrogen) gels. For co-localization of *DdP2X*-GFP and calmodulin, methanol-fixed *D. discoideum* cells were permeabilized with 0.1% Triton X-100 and blocked with 3% bovine serum albumin. Cells were incubated for 1 h at room temperature with mouse monoclonal anti-calmodulin primary antibody (1:20 dilution; Sigma), washed and subsequently incubated with Cy3-conjugated donkey anti-mouse secondary antibody (Jackson Laboratories). For localization of *DdP2X*-His in HEK 293 cells, the same protocol was performed but with anti-His tag (Invitrogen) as the primary antibody.

Generation of *DdP2X* receptor knockout. A 3-kilobase genomic fragment encoding *DdP2X* was amplified from wild-type AX4 genomic DNA and a tetracycline-blasticidin resistance cassette was inserted by *in vitro* transposition. This insertion is within exon 3 of the *p2xA* gene, about 650 base pairs downstream of the start codon. The linearized construct was transformed into *D. discoideum* followed by blasticidin selection. Homologous recombination was verified by polymerase chain reaction.

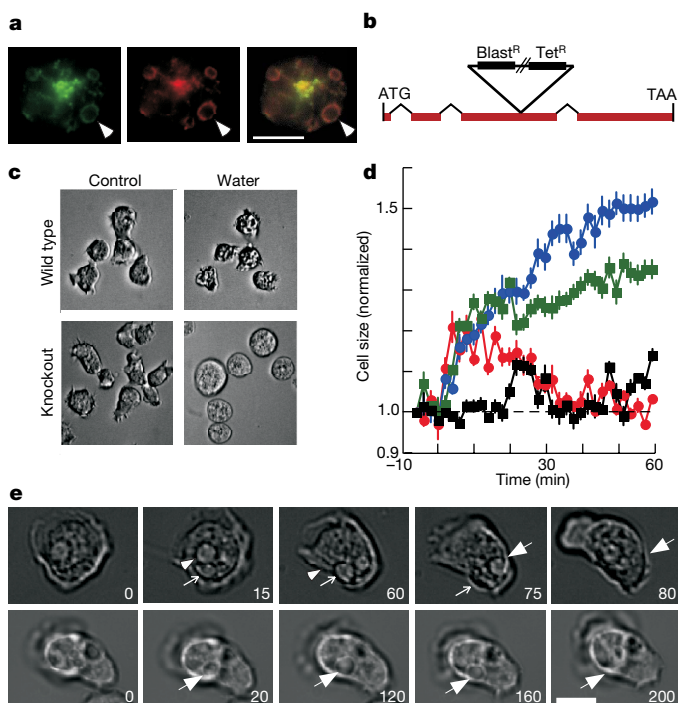


Figure 3 | *DdP2X* receptors are localized to the contractile vacuole and are required for cell volume regulation and contractile vacuole voiding. **a**, Co-localization of *DdP2X*-GFP (left, green) and calmodulin (middle, red), particularly on the membrane of the contractile vacuole (arrowheads indicate one such vacuole). Right, merged image. Scale bar, 10 μm . **b**, Structure of *D. discoideum* *p2xA* gene to illustrate position of cassette insertion. Disruption of the *p2xA* gene caused no observable differences in cell morphology or developmental cycle (data not shown). **c**, Bright-field micrographs of *D. discoideum* in HL5 growth medium, and 60 min after change from medium to distilled water. *DdP2X* mutant cells are swollen, whereas wild-type cells have recovered their normal volume. Scale bar, 10 μm . **d**, Time course of cell swelling and recovery. Wild-type cells (red circles) swell for 10–20 min and then regain their normal volume after about 30 min. *DdP2X* mutant cells (blue circles) continue to swell for 60 min. Wild-type cells treated with Cu^{2+} ions (10 μM ; green squares) continue to swell. Mutant cells expressing *DdP2X*-GFP (black squares) are also similar to wild-type cells except that they show less swelling in the first 10 min. Error bars indicate s.e.m. **e**, *DdP2X* receptors are required for normal contractile vacuole voiding. In wild-type amoebae (top row) the vacuoles (arrows) within the cell move to the membrane and void within 20–30 s. In *DdP2X*-disrupted amoebae (bottom row) this cycle is much prolonged: the vacuole identified takes almost 3 min to discharge. Numbers indicate time in seconds. Scale bar, 5 μm .

Received 28 November 2006; accepted 22 May 2007.

- North, R. A. Molecular physiology of P2X receptors. *Physiol. Rev.* **82**, 1013–1067 (2002).
- Khakh, B. S. & North, R. A. P2X receptors as cell-surface ATP sensors in health and disease. *Nature* **442**, 527–532 (2006).
- Gever, J. R., Cockayne, D. A., Dillon, M. P., Burnstock, G. & Ford, A. P. Pharmacology of P2X channels. *Pflügers Arch.* **452**, 513–537 (2006).
- Nicke, A. *et al.* P2X₁ and P2X₃ receptors form stable trimers: a novel structural motif of ligand-gated ion channels. *EMBO J.* **17**, 3016–3028 (1998).
- Barrera, N. P., Ormond, S. J., Henderson, R. M., Murrell-Lagnado, R. D. & Edwardson, J. M. Atomic force microscopy imaging demonstrates that P2X₂ receptors are trimers but that P2X₆ receptor subunits do not oligomerize. *J. Biol. Chem.* **280**, 10759–10765 (2005).
- Coddou, C. *et al.* Histidine 140 plays a key role in the inhibitory modulation of the P2X₄ nucleotide receptor by copper but not zinc. *J. Biol. Chem.* **278**, 36777–36785 (2003).
- Virginio, C., Church, D., North, R. A. & Surprenant, A. Effects of divalent cations, protons and calmidazolium at the rat P2X₇ receptor. *Neuropharmacology* **36**, 1285–1294 (1997).
- Roberts, J. A. *et al.* Molecular properties of P2X receptors. *Pflügers Arch.* **452**, 486–500 (2006).
- Vial, C., Roberts, J. A. & Evans, R. J. Molecular properties of ATP-gated P2X receptor ion channels. *Trends Pharmacol. Sci.* **25**, 487–493 (2004).
- Chaumont, S., Jiang, L. H., Penna, A., North, R. A. & Rassendren, F. Identification of a trafficking motif involved in the stabilization and polarization of P2X receptors. *J. Biol. Chem.* **279**, 29628–29638 (2004).
- Allen, R. D. The contractile vacuole and its membrane dynamics. *BioEssays* **22**, 1035–1042 (2000).
- Zhu, Q. & Clarke, M. Association of calmodulin and an unconventional myosin with the contractile vacuole complex of *Dictyostelium discoideum*. *J. Cell Biol.* **118**, 347–358 (1992).
- Heuser, J. Evidence for recycling of contractile vacuole membrane during osmoregulation in *Dictyostelium* amoebae. *Eur. J. Cell Biol.* **85**, 859–871 (2006).
- Allen, R. D. & Naitoh, Y. Osmoregulation and contractile vacuoles of protozoa. *Int. Rev. Cytol.* **215**, 351–394 (2002).
- Gerald, N. J., Sianom, M. & De Lozanne, A. The *Dictyostelium* LvsA protein is localized on the contractile vacuole and is required for osmoregulation. *Traffic* **3**, 50–60 (2002).
- Wu, W. I., Yajnik, J., Siano, M. & De Lozanne, A. Structure–function analysis of the BEACH protein LvsA. *Traffic* **5**, 346–355 (2004).
- Jentsch, T. J., Hubner, C. A. & Fuhrmann, J. C. Ion channels: function unravelled by dysfunction. *Nature Cell Biol.* **6**, 1039–1047 (2004).
- Masse, K., Eason, R., Bhamra, S., Dale, N. & Jones, E. A. Comparative genomic and expression analysis of the conserved NTPDase gene family in *Xenopus*. *Genomics* **87**, 366–381 (2006).
- Ramjeesingh, M., Huan, L. J., Garami, E. & Bear, C. E. Novel method for evaluation of the oligomeric structure of membrane proteins. *Biochem. J.* **342**, 119–123 (1999).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank H. Broomhead, L. Almond, K. Dossi and N. Aldren for their technical expertise during this study, and A. Mueller-Taubenberger for the gift of the pDEX 27 GFP vector. This work was supported by the Wellcome Trust, the Medical Research Council and the Lister Institute of Preventive Medicine.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to C.R.L.T. (christopher.thompson@manchester.ac.uk) or R.A.N. (r.a.north@manchester.ac.uk).

LETTERS

The menthol receptor TRPM8 is the principal detector of environmental cold

Diana M. Bautista^{1,2*}, Jan Siemens^{1,2*}, Joshua M. Glazer^{5*}, Pamela R. Tsuruda^{1,2}, Allan I. Basbaum^{3,4}, Cheryl L. Stucky⁵, Sven-Eric Jordt⁶ & David Julius^{1,2}

Sensory nerve fibres can detect changes in temperature over a remarkably wide range, a process that has been proposed to involve direct activation of thermosensitive excitatory transient receptor potential (TRP) ion channels^{1–4}. One such channel—TRP melastatin 8 (TRPM8) or cold and menthol receptor 1 (CMR1)—is activated by chemical cooling agents (such as menthol) or when ambient temperatures drop below ~26 °C, suggesting that it mediates the detection of cold thermal stimuli by primary afferent sensory neurons^{5,6}. However, some studies have questioned the contribution of TRPM8 to cold detection or proposed that other excitatory or inhibitory channels are more critical to this sensory modality *in vivo*^{7–10}. Here we show that cultured sensory neurons and intact sensory nerve fibres from TRPM8-deficient mice exhibit profoundly diminished responses to cold. These animals also show clear behavioural deficits in their ability to discriminate between cold and warm surfaces, or to respond to evaporative cooling. At the same time, TRPM8 mutant mice are not completely insensitive to cold as they avoid contact with surfaces below 10 °C, albeit with reduced efficiency. Thus, our findings demonstrate an essential and predominant role for TRPM8 in thermosensation over a wide range of cold temperatures, validating the hypothesis² that TRP channels are the principal sensors of thermal stimuli in the peripheral nervous system.

We generated TRPM8-deficient mice through targeted deletion of a genomic region encoding amino acids 594–661 within the presumptive cytoplasmic amino-terminal domain. In addition to removing coding information, this manipulation introduced a stop codon before and a frameshift after the deleted segment (Fig. 1a). Successful targeting was verified by Southern blotting and RT-PCR analysis of transcripts from trigeminal ganglia of wild-type and mutant mice (Fig. 1b). The resulting *TRPM8*^{−/−} mice were normal in overall appearance and viability, and matings between heterozygous animals produced siblings with normal mendelian distributions for gender and genotype. Moreover, mutant and wild-type littermates showed no differences in core body temperature (37.4 ± 0.15 °C and 37.4 ± 0.13 °C, respectively; *n* = 11 mice per genotype).

Sensory ganglia from TRPM8-deficient mice appeared anatomically normal, as was assessed by staining with a variety of immunological probes that label peptidergic, unmyelinated or vanilloid-responsive neurons (Fig. 1c, d). Thus, antibodies directed against substance P (peripherin) or the capsaicin receptor (TRPV1) revealed an identical prevalence of labelled cells in trigeminal ganglia from wild-type and mutant mice. In contrast, TRPM8 immunoreactivity was specifically absent from trigeminal ganglia or the peripheral free nerve endings of adult mutant mice. Central projections of primary afferent fibres to the spinal cord of *TRPM8*^{−/−} mice were also devoid of TRPM8

immunoreactivity, while those from wild-type littermates showed a robust pattern of staining that was confined to the most superficial layer of the dorsal horn (lamina I), a region that is innervated by primary afferent nociceptors, including those expressing TRPV1. Importantly, polymerase chain reaction (PCR) analysis showed that *TRPM8*^{−/−} mice produce a truncated non-functional transcript from the disrupted locus, enabling us to demonstrate by *in situ* hybridization histochemistry that loss of TRPM8 protein expression does not eliminate neurons that normally express this channel (Supplementary Fig. 1).

Functional disruption of the *TRPM8* gene was assessed by using live-cell calcium imaging to measure the sensitivity of trigeminal and dorsal root ganglion neurons to cooling agents. Consistent with previous observations^{5,11}, approximately 17% of neurons cultured from wild-type ganglia showed rapid and robust responses to bath applied menthol (100 μM) or the super-cooling agent icilin (100 μM). Neurons from *TRPM8*^{−/−} mice were completely unresponsive to either compound (Fig. 2a, b). In contrast, TRPM8-deficient neurons displayed normal sensitivity to capsaicin or the TRPA1 agonist, allyl isothiocyanate (mustard oil)^{12,13} (Fig. 2b). Icilin, which is a weak TRPA1 agonist¹⁴, activated all mustard-oil-sensitive neurons at concentrations over 500 μM, regardless of genotype (not shown). Together, these results verify selective loss of TRPM8 activity in mutant mice and suggest that this channel is the sole target through which menthol and icilin mediate their cooling effects.

Next, we asked whether TRPM8 ablation affects cellular cold sensitivity. When challenged with a cooling gradient (30 to 8 °C over 30 s), ~22% of trigeminal neurons from wild-type mice showed significant increases in cellular calcium levels, as previously described¹¹ (Fig. 2c, d). These cold-sensitive neurons could be segregated into two distinct populations: most (77%) were menthol-sensitive, showed a rapid activation rate ($dCa^{2+}/dT = 40 \text{ nmol s}^{-1}$), and had an average response threshold of 22 °C when cooled from a holding temperature of 30 °C. As observed with the cloned TRPM8 channel, the activation threshold of this neuronal subpopulation shifted to 16 °C when cooled from a holding temperature of 22 °C (not shown), suggesting that both cloned and native channels show adaptation¹⁵. The remaining 23% of cold-sensitive neurons (5% of all neurons) were menthol-, capsaicin- and mustard-oil-insensitive; they showed a substantially slower cold activation rate ($dCa^{2+}/dT = 4 \text{ nmol s}^{-1}$), smaller peak cold-evoked calcium response (300 nM), and had an average response threshold of 12 °C irrespective of starting temperature. Trigeminal neurons from *TRPM8*^{−/−} mice showed a dramatic reduction in the total number of cold-sensitive neurons, with only 4% of cells responding to the cold ramp (Fig. 2c, d). The response properties of these residual cold-sensitive neurons closely matched those of the

¹Department of Physiology, ²Department of Cellular and Molecular Pharmacology, ³Departments of Anatomy and Physiology, ⁴W. M. Keck Center for Integrative Neuroscience, University of California, San Francisco, California 94143, USA. ⁵Department of Cell Biology, Neurobiology and Anatomy, Medical College of Wisconsin, Milwaukee, Wisconsin 53226, USA. ⁶Department of Pharmacology, Yale University School of Medicine, New Haven, Connecticut 06520, USA.

*These authors contributed equally to this work.

menthol-insensitive population present in cultures from wild-type ganglia. Thus *TRPM8* ablation results in a profound reduction in both the number and magnitude of cellular cold responses, leaving intact a small cohort of neurons that respond to cold with an activation threshold of $\sim 12^{\circ}\text{C}$, and probably correspond to the menthol-insensitive component of cold-sensitive neurons from wild-type ganglia. No significant differences in chemical or thermal responses were observed between dorsal root ganglia and trigeminal neurons, irrespective of genotype (Supplementary Table 1).

In contrast to differences in cold sensitivity, *TRPM8*-deficient neurons showed normal responses to heat. When exposed to a heat ramp (30 to 45°C in 45 s), 44% of trigeminal neurons from *TRPM8*^{-/-} mice were heat-sensitive, compared with 41% from wild-type littermates. Conversely, mice lacking the capsaicin receptor TRPV1 showed a normal prevalence of cold-sensitive neurons, but a substantial reduction in the number of cells responding to heat, as expected (Fig. 2d).

We next used the skin-nerve preparation to ask whether *TRPM8* mediates cold responses at the peripheral terminals of cutaneous afferent fibres^{16,17}. When subjected to a cold ramp (32 to 2°C over 20 s), 35% of C fibres from wild-type mice showed robust action potential firing. As expected, cold-evoked firing rates of these fibres showed adaptation within the first 8 s of the cold ramp (Supplementary Fig. 2b). In contrast, only 5.4% of C fibres from

TRPM8-deficient mice responded to cold ($P < 0.0001$; Fisher's exact test; Fig. 3a, b), and these residual cold-sensitive units had significantly decreased firing rates (Fig. 3b; $P < 0.0001$; *t*-test). Basal (ongoing) action potential firing at skin temperature (32°C) is a property of cold-sensitive fibres^{18,19}, and indeed we found that wild-type cold-activated C fibres had significantly higher baseline firing rates than cold-insensitive fibres (Supplementary Fig. 2a). The number of C fibres showing baseline activity did not differ significantly when comparing wild-type and *TRPM8*-deficient preparations (60% versus 45%; $P = 0.14$, Fisher's exact test). However, C fibres from *TRPM8*-deficient mice showed a significantly reduced basal firing rate, akin to that observed in cold-insensitive units from wild-type littermates (Supplementary Fig. 2a). These results suggest that *TRPM8* also modulates the characteristic basal firing rate of the majority of cold-sensitive C fibres. Among myelinated A δ fibres, cold also elicited firing of A-mechanoreceptors (AM), albeit less robustly than generally observed with unmyelinated C fibres (Fig. 3c). In wild-type preparations, 16.7% of these units showed cold-evoked action potentials, whereas only 1 of 26 (3.9%) AM fibres from *TRPM8*-deficient mice responded with comparable firing rates (Fig. 3d).

In contrast to these robust deficits in cold sensitivity, the electrical response characteristics of *TRPM8*-deficient fibres, including conduction velocity and electrical activation threshold, were almost identical to those of wild-type littermates. Furthermore, the

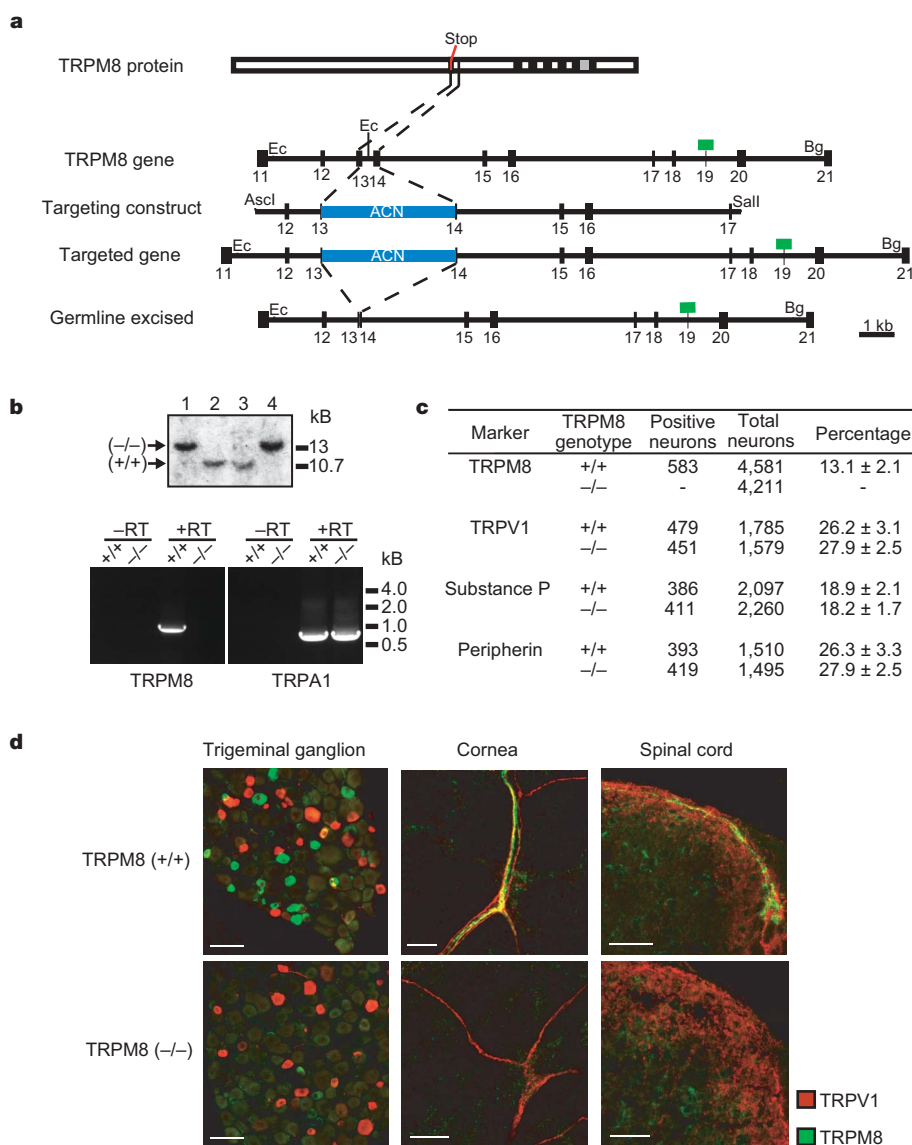


Figure 1 | Generation and histological analysis of *TRPM8*-deficient mice. **a**, *TRPM8* gene targeting strategy introduces a stop codon within the protein-coding region. Transmembrane and pore loop domains are indicated by black and grey bars, respectively; exons are numbered (black bars on genomic map); blue bar depicts self-excising ACN neomycin expression cassette; green bar denotes 3' probe. Restriction sites indicated are *EcoRI* (Ec), *BglI* (Bg), *Sall* and *Ascl*. **b**, Southern blot (top) confirms targeting of *TRPM8* locus, where *BglI* and *EcoRI* digests produce expected fragments of 13 and 10.7 kilobases (kb) for wild-type and mutant alleles, respectively. RT-PCR (bottom) confirms the absence of normal *TRPM8* transcripts in trigeminal ganglia from *TRPM8*^{-/-} mice. Primers amplifying a 1 kb region from exon 13 of the *TRPM8* gene do not generate a product with RNA from *TRPM8*^{-/-} ganglia. As positive control, a 700 base pair (bp) region of *TRPA1* was amplified from wild-type or *TRPM8*^{-/-} ganglia. **c**, Immunohistochemical analysis using anti-TRPV1, anti-substance P, or anti-peripherin antibodies shows similar distribution of labelled cells in *TRPM8*-deficient and wild-type trigeminal ganglia. In contrast, *TRPM8* immunoreactivity is absent from *TRPM8*^{-/-} ganglia, whereas wild-type sections show staining in $\sim 13\%$ of neurons (2–3 animals per genotype). Percentages represent mean \pm s.e.m. **d**, Immunostaining of trigeminal ganglia (left), corneal afferents (middle), and spinal cord dorsal horn (right) with anti-TRPM8 (green) and anti-TRPV1 (red) antibodies reveals selective loss of *TRPM8* expression in *TRPM8*^{-/-} mice. Scale bars, 50 μm .

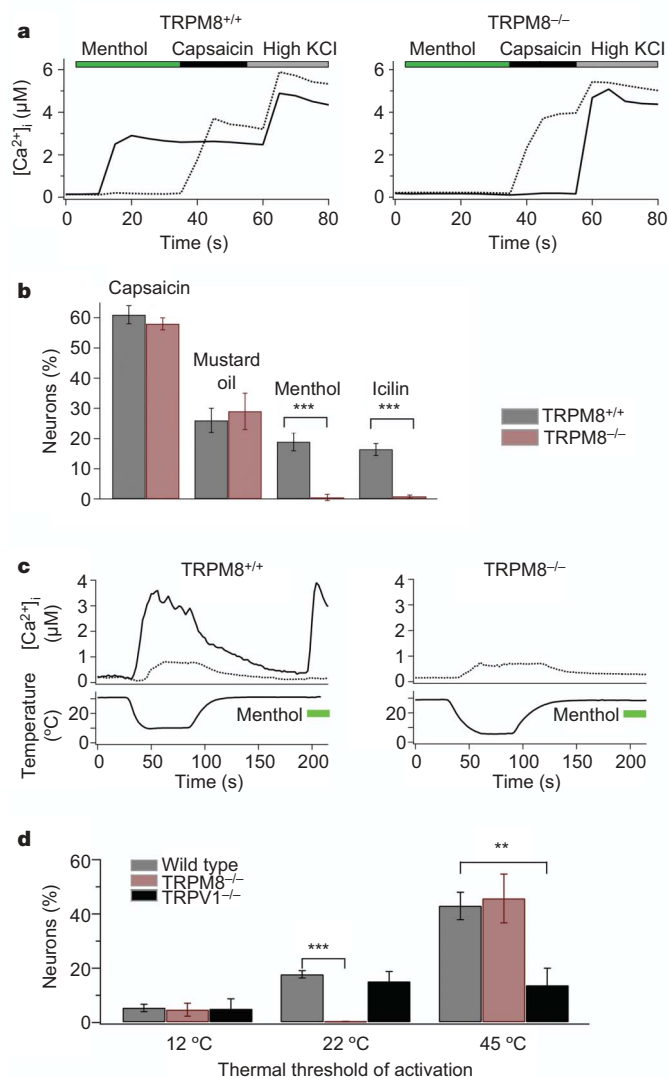


Figure 2 | TRPM8-deficient neurons show loss of menthol and cold sensitivity. **a**, Trigeminal neurons from TRPM8^{+/+} (left) or TRPM8^{-/-} (right) mice were challenged with menthol (100 μ M), followed by capsaicin (1 μ M), then high KCl (75 mM). Responses were assessed by calcium imaging. TRPM8^{+/+} ganglia displayed two distinct subpopulations: menthol-sensitive (solid trace) and capsaicin-sensitive (dotted trace) neurons. No menthol sensitivity was detected in TRPM8^{-/-} neurons ($n = 50$ cells per trace). Similar results were obtained with 50 or 250 μ M menthol (not shown). **b**, Prevalence of sensory neurons responding to capsaicin (1 μ M), allyl isothiocyanate (mustard oil, 100 μ M), menthol (100 μ M) or icilin (100 μ M) from TRPM8^{+/+} and TRPM8^{-/-} ganglia. **c**, Trigeminal neurons were challenged with cold and menthol (green bar) and responses assessed by calcium imaging. In wild-type ganglia, menthol-sensitive neurons (solid trace) showed robust cold responses (threshold 22 $^{\circ}$ C \pm 1.5 $^{\circ}$ C), and some menthol-insensitive neurons (dotted trace) showed low-level responses (threshold 12 $^{\circ}$ C \pm 1.5 $^{\circ}$ C). TRPM8^{-/-} ganglia contained few cold responders, which resembled the latter population. No significant differences were observed between dorsal root ganglia and trigeminal neurons, irrespective of genotype (Supplementary Table 1). **d**, Comparison of cold and heat sensitivity for wild type, TRPM8^{-/-} and TRPV1^{-/-} neurons. For wild-type ganglia (grey bars) 4% of neurons were activated with a threshold of 12 $^{\circ}$ C, 17% at 22 $^{\circ}$ C, and 44% at 45 $^{\circ}$ C. The 22 $^{\circ}$ C threshold group was absent from TRPM8-deficient ganglia (red bars), while the 12 $^{\circ}$ C and 45 $^{\circ}$ C populations were unaltered. In contrast, the 45 $^{\circ}$ C population was substantially reduced (to 14%) in TRPV1-deficient ganglia (black bars), whereas cold-sensitive populations were unaltered ($n > 420$ neurons per genotype). ** $P < 0.01$, *** $P < 0.001$, one-way analysis of variance (ANOVA) analysis followed by Tukey's HSD post-hoc analysis. No significant differences were observed when comparing chemical or thermal sensitivity of wild-type and TRPM8 heterozygous neurons (not shown). Graphs display mean \pm s.e.m.

mechanical thresholds of C and AM fibres did not differ significantly between genotypes (Supplementary Fig. 2d). Taken together, these findings demonstrate that loss of TRPM8 expression results in a selective deficit in cold sensitivity of both C and AM fibres, without affecting other aspects of electrical and mechanical excitability.

To assess behavioural cold sensitivity, we first measured acute flinching responses to acetone-evoked evaporative cooling of the hind paw. TRPM8-deficient mice showed a significant reduction ($\sim 50\%$) in this nocifensive response compared to wild-type littermates (Fig. 4a). Next, we asked whether these animals also exhibit deficits in their ability to discriminate between warm and cold

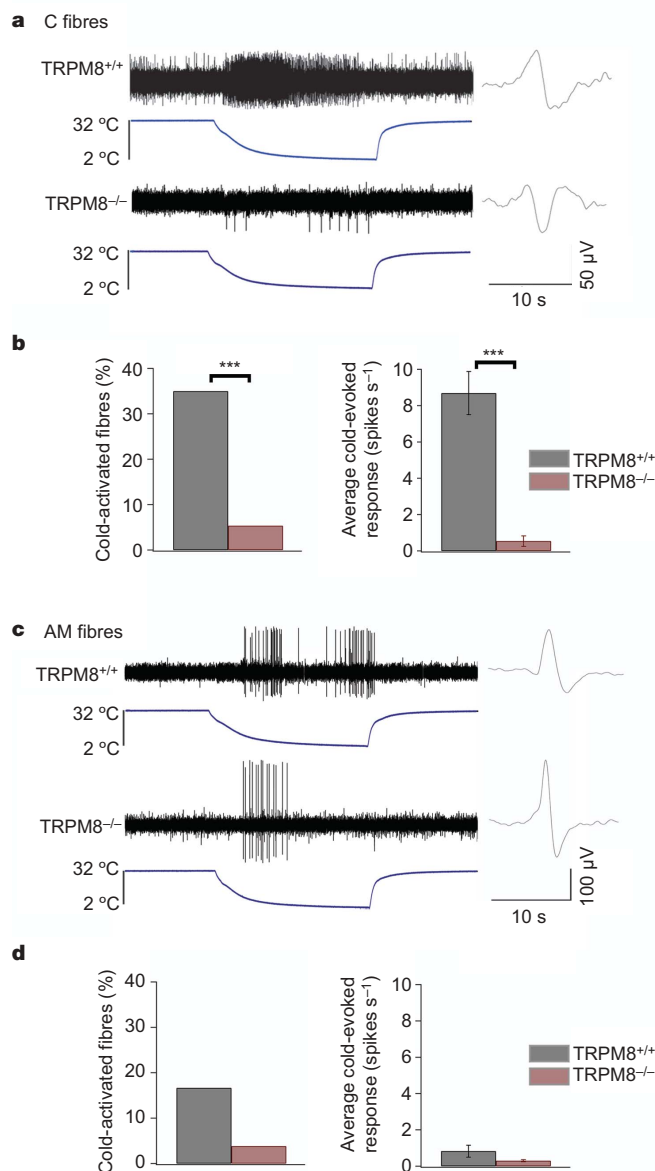


Figure 3 | TRPM8-deficient nerve fibres show loss of cold sensitivity. **a**, Typical response of wild-type (top) and TRPM8-deficient (bottom) cutaneous C fibre to a cold ramp (32 to 2 $^{\circ}$ C, over 20 s). Action potential waveform is shown to right of trace. **b**, Percentage of C fibres responding to cold ramp in wild-type (black; $n = 60$) versus TRPM8^{-/-} (red; $n = 56$) mice (left). Average cold-evoked action potential firing rate in C fibres of wild-type ($n = 21$) versus TRPM8^{-/-} ($n = 3$) mice (right; mean \pm s.e.m.). **c**, Typical response of wild-type (top) and TRPM8-deficient (bottom) cutaneous AM fibre to a cold ramp (32 to 2 $^{\circ}$ C, over 20 s). **d**, Percentage of AM fibres responding to cold ramp in wild-type (black; $n = 24$) versus TRPM8^{-/-} (red; $n = 26$) mice (left). Average cold-evoked action potential firing rate in AM fibres of wild-type ($n = 4$; mean \pm s.e.m.) versus TRPM8^{-/-} ($n = 1$) mice (right). *** $P < 0.001$, Student's t -test or Fisher's exact test.

surfaces. Animals were allowed to explore adjacent surfaces, with one held at 30 °C and the other ranging from 49 to 5 °C (Fig. 4b). The percentage of time spent on the 30 °C surface was measured over a 5 min period. When placed on equivalent temperatures (both 30 °C), neither wild-type nor mutant mice displayed a preference, spending equal time on each side, as expected (Fig. 4c). As the variable plate was cooled, wild-type mice showed a clear preference for the 30 °C side, spending less time on the cold side as temperature decreased. In striking contrast, TRPM8-deficient littermates showed relatively little, if any, preference for the 30 °C side until the adjacent surface dropped below 15 °C (Fig. 4c), a threshold that has been suggested to represent a demarcation between innocuous cool and noxious cold in primates²⁰. But even at these lower temperatures, TRPM8^{-/-} mice explored the cold surface for an extended period, taking substantially more time than wild-type littermates to appreciate the aversive nature of the environment relative to the warmer side. Wild-type and mutant mice also exhibited striking differences in their patterns of

exploratory behaviour when approaching a cold surface. After their initial contact with the colder surface, wild-type animals approached the plate with apparent trepidation, making only rapid and brief contact with their front paws before withdrawing to the warmer side. In contrast, TRPM8-deficient mice made full transitions to the cold side and showed relatively uninhibited exploratory behaviour, as reflected by the greater number of crossings between adjacent surfaces (Supplementary Video 1). Nonetheless, mutant mice displayed normal preference for 30 °C over noxious heat (49 °C), and thus did not exhibit a general deficiency in thermosensation (Fig. 4c).

We also examined nocifensive behaviours by placing animals on a single plate set to a variety of noxious temperatures. Equivalent licking, flinching or shivering responses were observed when TRPM8-deficient or wild-type mice were placed on a plate set to 52, 10, 0 or -5 °C (Fig. 4d). No behavioural responses were observed at non-noxious settings of 20 and 30 °C (not shown). Similar findings were obtained using a behavioural paradigm that measures time to paw withdrawal from a radiant heat source, where mutant and wild-type mice showed similar latencies (11.8 ± 0.42 s and 12.2 ± 0.49 s, respectively; $n = 11$ per genotype; Supplementary Fig. 3). Baseline mechanical sensitivity was also normal in TRPM8-deficient mice; paw-withdrawal thresholds to graded mechanical force did not differ with genotype (0.77 ± 0.12 g and 0.73 ± 0.09 g for TRPM8^{-/-} and TRPM8^{+/+} littermates, respectively; $n = 23$ per genotype; Supplementary Fig. 3). Taken together, our results show that TRPM8 is specifically required for normal cold sensation, but not for the detection of pressure or heat.

Cloning of the capsaicin receptor TRPV1 and its heat-sensitive homologue TRPV2 led to the prediction that a family of TRP channel subtypes mediates peripheral temperature detection over a wide physiological range, from hot to cold². Molecular characterization of TRPM8 bolstered this hypothesis because the electrophysiological and pharmacological properties of the cloned channel resemble those of a major subclass of native cold-evoked membrane currents^{5,10,15,21}. However, some studies have suggested that a significant component of cold-evoked neuronal depolarization or fibre excitation is mediated by other entities, such as epithelial sodium channels, TRPA1, background potassium channels, or ion pumps^{7,8,14,22–24}. It has also been proposed that cold evokes psychophysical sensations through its effects on vascular tone, which are then transduced to sensory afferents indirectly^{25,26}. Our present results show that TRPM8 is essential for a large component of cold sensitivity, whether assessed at the cellular, nerve terminal or behavioural level. These findings solidify the importance of this channel, and the primary afferent nerve fibre, for the direct detection of environmental cold. More generally, our results confirm the hypothesis that TRP channels serve as the principal transducers of thermal stimuli within the mammalian peripheral nervous system.

TRPM8-deficient mice retain a small, but finite number of cold-sensitive neurons, consistent with our fibre recordings showing a small residual population of cold-sensitive C and AM units. Moreover, these animals display aversion to cold at temperatures below 15 °C, which is considered to be an approximate demarcation between innocuous cool and noxious cold^{20,27}. These observations suggest that TRPM8-independent mechanisms exist for the detection of cold, and are consistent with a number of cellular studies describing a population of menthol-insensitive, cold-sensitive neurons^{11,14,28}. The residual cold-sensitive neurons and fibres that we observe in TRPM8-deficient mice probably correspond to this population on the basis of the similar prevalence and activation threshold of ~12 °C. The molecular mechanism underlying this small, but finite cold sensitivity remains unknown, but could involve modulation of other excitatory or inhibitory channels on primary afferent sensory neurons. The mustard-oil receptor TRPA1 has been proposed to function as a detector of noxious cold by menthol-insensitive nociceptors¹⁴, although recent electrophysiological and genetic studies suggest otherwise^{11,15,29,30}. In any case, our present

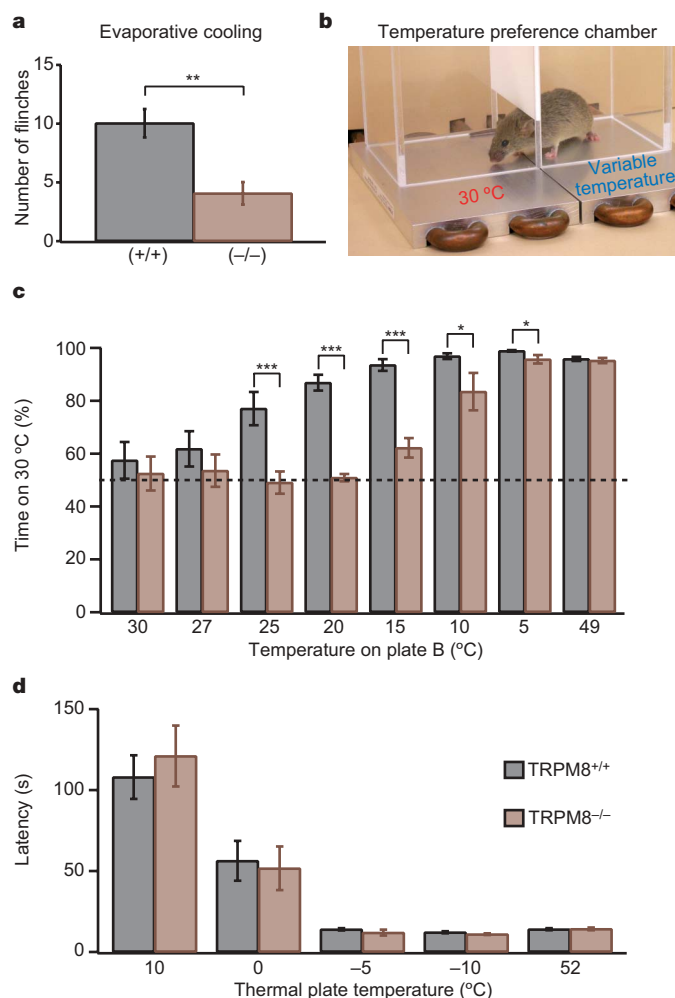


Figure 4 | TRPM8 is required for normal cold-evoked behaviours.

a, Licking and flinching in response to evaporative cooling was measured for 1 min following application of acetone (50 μ l) to the hind paw. TRPM8^{-/-} mice displayed significantly decreased behaviour compared to TRPM8^{+/+} littermates ($n = 21$ mice per genotype). **b**, Mouse in temperature preference chamber. **c**, TRPM8^{+/+} and TRPM8^{-/-} littermates were allowed to choose between adjacent surfaces adjusted to 30 °C versus a range of temperatures, as shown. The percentage of time spent at 30 °C over a 5 min period is shown ($n \geq 10$ mice per genotype). **d**, Latency to licking or flinching of the hind paw was measured following placement of mice on a single platform set to a variety of temperatures, as shown. TRPM8^{+/+} and TRPM8^{-/-} littermates displayed similar latencies at all temperatures examined ($n = 6$ mice per genotype). * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, one-way ANOVA with Tukey's HSD post-hoc analysis. Graphs display mean \pm s.e.m.

results show that the residual cold-sensitive neurons in TRPM8-deficient mice are not mustard-oil-sensitive, suggesting that TRPA1 does not underlie this response. Consistent with these cellular findings, TRPA1-deficient mice (males or females) were indistinguishable from wild-type littermates when examined for behavioural responses to cold using the two-plate choice test (Supplementary Fig. 4). Furthermore, a recent analysis of skin-nerve preparations from TRPA1-deficient mice showed no deficit in cold responsiveness in either C or A fibre nociceptors (C.L.S. and personal communication from K. Y. Kwan).

Our behavioural analysis shows that TRPM8 constitutes the primary, if not sole determinant of thermosensitivity in the cool range, and that it also contributes to the detection of noxious cold, defined as temperatures below 15 °C. It remains to be determined whether the small residual population of cold-sensitive neurons and fibres enables TRPM8^{-/-} mice to recognize noxious cold. In addition to activating TRPM8, noxious cold may also activate sensory nerve fibres indirectly by provoking tissue injury or changes in vascular tone, thereby leading to the production or release of pro-algesic agents that evoke pain by stimulating nociceptors. Alternatively, aversion to noxious cold may be triggered by the inhibition of warm-sensing fibres. TRPM8-deficient mice now provide a genetic model with which to dissect this residual aspect of cold sensation.

METHODS

Mouse genetics. TRPM8-deficient mice were generated through homologous recombination in E14Tg2A.4 embryonic stem cells. Positive clones were injected into C57/Bl6 blastocysts and male chimaeras mated to C57/Bl6 females. Heterozygotes thus produced were crossed to generate paired F2 littermates for all studies.

Histology. Frozen paraformaldehyde-fixed tissue sections were prepared from trigeminal ganglia of wild-type or TRPM8-deficient mice and incubated with a variety of previously characterized antibody or nucleic acid probes. For any given marker, quantification of labelled cells was assessed while blind to genotype.

Cell and nerve fibre physiology. Dissociated sensory neurons from trigeminal or dorsal root ganglia of wild-type and TRPM8-deficient mice were placed in culture and analysed for chemical or thermal sensitivity using Fura 2-AM based ratiometric calcium imaging. The skin-nerve preparation was used to obtain electrophysiological recordings from desheathed saphenous nerve. Single afferent units were identified using an unbiased electrical search stimulus and subsequently examined for responsiveness to a cold ramp in which perfusate temperature decreased from 32 to 2 °C over 20 s. Both calcium imaging and nerve fibre recording were carried out with preparations from individual littermates while blind to genotype.

Behaviour. Cold sensitivity was assessed using acetone-evoked evaporative cooling or two-plate choice tests. Response latencies to noxious temperatures were measured using radiant heating or single hot/cold plate paradigms. Mechanical thresholds were determined using calibrated von Frey filaments. Behavioural studies were carried out with paired littermates while blind to genotype. All experiments were performed according to the policies and recommendations of the International Association for the Study of Pain and approved by the University of California, San Francisco Institution Animal Care and Use Committee.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 22 March; accepted 14 May 2007.

Published online 30 May 2007.

- Jordt, S. E., McKemy, D. D. & Julius, D. Lessons from peppers and peppermint: the molecular logic of thermosensation. *Curr. Opin. Neurobiol.* **13**, 487–492 (2003).
- Caterina, M. J., Rosen, T. A., Tominaga, M., Brake, A. J. & Julius, D. A capsaicin-receptor homologue with a high threshold for noxious heat. *Nature* **398**, 436–441 (1999).
- Dhaka, A., Viswanath, V. & Patapoutian, A. Trp ion channels and temperature sensation. *Annu. Rev. Neurosci.* **29**, 135–161 (2006).
- Lumpkin, E. A. & Caterina, M. J. Mechanisms of sensory transduction in the skin. *Nature* **445**, 858–865 (2007).

- McKemy, D. D., Neuhauser, W. M. & Julius, D. Identification of a cold receptor reveals a general role for TRP channels in thermosensation. *Nature* **416**, 52–58 (2002).
- Peier, A. M. *et al.* A TRP channel that senses cold stimuli and menthol. *Cell* **108**, 705–715 (2002).
- Viana, F., de la Pena, E. & Belmonte, C. Specificity of cold thermotransduction is determined by differential ionic channel expression. *Nature Neurosci.* **5**, 254–260 (2002).
- Madrid, R. *et al.* Contribution of TRPM8 channels to cold transduction in primary sensory neurons and peripheral nerve terminals. *J. Neurosci.* **26**, 12512–12525 (2006).
- Reid, G. & Flonta, M. Cold transduction by inhibition of a background potassium conductance in rat primary sensory neurones. *Neurosci. Lett.* **297**, 171–174 (2001).
- McKemy, D. D. How cold is it? TRPM8 and TRPA1 in the molecular logic of cold sensation. *Mol. Pain* **1**, 16–22 (2005).
- Bautista, D. M. *et al.* TRPA1 mediates the inflammatory actions of environmental irritants and proalgesic agents. *Cell* **124**, 1269–1282 (2006).
- Bandell, M. *et al.* Noxious cold ion channel TRPA1 is activated by pungent compounds and bradykinin. *Neuron* **41**, 849–857 (2004).
- Jordt, S. E. *et al.* Mustard oils and cannabinoids excite sensory nerve fibres through the TRP channel ANKTM1. *Nature* **427**, 260–265 (2004).
- Story, G. M. *et al.* ANKTM1, a TRP-like channel expressed in nociceptive neurons, is activated by cold temperatures. *Cell* **112**, 819–829 (2003).
- Reid, G. ThermoTRP channels and cold sensing: what are they really up to? *Pflugers Arch.* **451**, 250–263 (2005).
- Stucky, C. L. *et al.* Overexpression of nerve growth factor in skin selectively affects the survival and functional properties of nociceptors. *J. Neurosci.* **19**, 8509–8516 (1999).
- Koltzenburg, M., Stucky, C. L. & Lewin, G. R. Receptive properties of mouse sensory neurons innervating hairy skin. *J. Neurophysiol.* **78**, 1841–1850 (1997).
- Iggo, A. Cutaneous thermoreceptors in primates and sub-primates. *J. Physiol. Lond.* **200**, 403–430 (1969).
- Schafer, K., Braun, H. A. & Kurten, L. Analysis of cold and warm receptor activity in vampire bats and mice. *Pflugers Arch.* **412**, 188–194 (1988).
- Rainville, P., Chen, C. C. & Bushnell, M. C. Psychophysical study of noxious and innocuous cold discrimination in monkey. *Exp. Brain Res.* **125**, 28–34 (1999).
- Reid, G. & Flonta, M. L. Cold current in thermoreceptive neurons. *Nature* **413**, 480 (2001).
- Maingret, F. *et al.* TREK-1 is a heat-activated background K(+) channel. *EMBO J.* **19**, 2483–2491 (2000).
- Askwith, C. C., Benson, C. J., Welsh, M. J. & Snyder, P. M. DEG/ENAC ion channels involved in sensory transduction are modulated by cold temperature. *Proc. Natl Acad. Sci. USA* **98**, 6459–6463 (2001).
- Pierau, F. K., Torrey, P. & Carpenter, D. O. Mammalian cold receptor afferents: role of an electrogenic sodium pump in sensory transduction. *Brain Res.* **73**, 156–160 (1974).
- Kreh, A., Anton, F., Gilly, H. & Handwerker, H. O. Vascular reactions correlated with pain due to cold. *Exp. Neurol.* **85**, 533–546 (1984).
- Fruhstorfer, H. & Lindblom, U. Vascular participation in deep cold pain. *Pain* **17**, 235–241 (1983).
- Hensel, H. & Zotterman, Y. The effect of menthol on the thermoreceptors. *Acta Physiol. Scand.* **24**, 27–34 (1951).
- Babes, A., Zorzon, D. & Reid, G. A novel type of cold-sensitive neuron in rat dorsal root ganglia with rapid adaptation to cooling stimuli. *Eur. J. Neurosci.* **24**, 691–698 (2006).
- Zurborg, S., Yurgionas, B., Jira, J. A., Caspani, O. & Heppenstall, P. A. Direct activation of the ion channel TRPA1 by Ca(2+). *Nature Neurosci.* **10**, 277–279 (2007).
- Doerner, J. F., Gisselmann, G., Hatt, H. & Wetzel, C. H. Transient receptor potential channel A1 is directly gated by calcium ions. *J. Biol. Chem.* (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank R. Nicoll for discussion and criticism, M. Tominaga for providing TRPM8 antiserum, T. Nikai for advice regarding behavioural assays, and J. Poblete for technical assistance throughout. This work was supported by grants from the NIH (D.J., A.I.B., S.-E.J. and C.L.S.), a Burroughs Wellcome Fund Career Award in Biomedical Sciences (D.M.B.) and a postdoctoral fellowship from the International Human Frontier Science Program Organization (J.S.).

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to D.J. (julius@cmp.ucsf.edu), S.-E.J. (sven.jordt@yale.edu) or C.L.S. (cstucky@mcw.edu).

METHODS

TRPM8 gene disruption. The targeting construct deleted 569 base pairs of genomic sequence within exons 13 and 14. Genomic DNA flanking the deleted region (1.9 and 8.7 kb) was PCR-amplified with KOD polymerase (Novagen/Toboyo) from clones 326P3 and 572M11 of the RPCI-22 Bac library (Invitrogen), subcloned and sequenced. Fragments were transferred to pACN, a vector with a neomycin cassette that is removed through self-excision in the male germline³¹. Linearized targeting construct was transfected into E14Tg2A.4 mouse embryonic stem cells (UCSF Core Facility). G418-resistant clones were screened for homologous recombination by PCR and verified by Southern blotting using 5' and 3' flanking probes as well as a neomycin cassette probe. Targeted embryonic stem cell clones were injected into C57/Bl6 blastocysts and chimaeric mice mated to C57/Bl6 females. Heterozygotes were mated to produce paired littermates for all studies. Germline transmission of the mutated allele and excision of the selection cassette were verified by PCR analysis and Southern blotting.

Histology and PCR. Cryostat sections (15 µm thick) were prepared from paraformaldehyde-fixed trigeminal ganglia as described². Sections were incubated overnight at 4 °C with affinity purified guinea pig anti-mouse TRPV1 antiserum diluted 1:1,000 (Julius laboratory), affinity purified rabbit anti-rat TRPM8 antiserum diluted 1:1,000 (kindly provided by M. Tominaga), guinea pig anti-substance P antiserum diluted 1:20,000 (kindly provided by J. Maggio), or goat anti-peripherin antiserum (Santa Cruz Biotechnology). Primary antisera were visualized by subsequent incubation with AlexaFluor 594- or 488-coupled secondary antibody (Invitrogen). For RT-PCR, total RNA was isolated from trigeminal ganglia of TRPM8^{-/-} or TRPM8^{+/+} mice and 1 µg was used to synthesize first-strand complementary DNA using SuperScript II reverse transcriptase (Invitrogen) according to the manufacturer's protocol. PCR was carried out using primer pairs corresponding to nucleotides 1909–1933 and 2907–2931 for mouse TRPM8, and 1681–1710 and 2453–2481 for mouse TRPA1.

Calcium imaging. Dissociation and culturing of mouse trigeminal or dorsal root ganglion neurons and ratiometric calcium imaging were carried out as previously described¹¹. For calcium imaging experiments, analysis was performed on neurons cultured from individual littermates while blind to genotype. Thermal threshold were defined as the temperature at which the response exceeds the standard deviation of the baseline by fivefold.

Skin-nerve preparation and fibre analysis. Mice were anaesthetized with isoflurane and sacrificed via cervical dislocation. Saphenous nerve and skin from the medial dorsum of the hind paw were dissected free and placed corium side up into a tissue bath superfused with oxygen-saturated synthetic interstitial fluid containing (in mM): 123 NaCl, 3.5 KCl, 0.7 MgSO₄, 1.7 NaH₂PO₄, 2.0 CaCl₂, 9.5 sodium gluconate, 5.5 glucose, 7.5 sucrose and 10 HEPES, 290 mOsm, pH 7.45 ± 0.05, at 32.0 ± 0.5 °C. The saphenous nerve was desheathed and teased into fine filaments for extracellular recordings as previously described^{16,17}. Single afferent units were identified using an electrical search stimulus in which a Teflon-coated steel needle electrode (2 MΩ impedance, uninsulated tip diameter of 10 µm) was inserted systematically across the skin preparation while square-wave pulses (500 µs, 6 mA) were applied. Following action potential (signal-to-noise ratio > 3) identification, the surrounding tissue was probed to identify the lowest electrical threshold. This point corresponded reliably to the most mechanically sensitive part of the receptive field in mechanically sensitive fibres as determined by calibrated von Frey filaments. Conduction velocity was calculated as previously described¹⁶. Units conducting slower than 1.2 m s⁻¹ were classified as unmyelinated C fibres, those conducting between 1.2 and 10 m s⁻¹ were classified as thinly myelinated Aδ fibres, and those exhibiting slowly adapting responses to sustained force were further classified as AM fibres. C fibres and AM fibres were characterized for cold responsiveness. Cold stimuli were delivered as follows: first, physiological zero buffer (32 °C) was applied to the isolated receptive field for 20 s. Next, a cold ramp was applied from 32 to 2 °C over a 20 s duration. Action potentials were recorded and analysed as previously described¹⁶. The number of action potentials evoked during a 32 °C buffer stimulus (20 s) was subtracted from the total number evoked during the cold ramp. For each fibre, the cold threshold was identified as the temperature at which the firing rate exceeded five times the standard deviation of the baseline firing rate sampled during the 32 °C stimulus. All experiments and analyses were performed blind to genotype. One-way ANOVA and Student's *t*-test were used to determine significance between wild-type and mutant responses, as indicated in the figure legends.

Behaviour. All mice (20–35 g) were housed with 12 h light–dark cycle at an ambient temperature of 21 °C and experiments were performed under the policies and recommendations of the International Association for the Study of Pain and approved by the University of California, San Francisco Institution Animal Care and Use Committee. Core body temperature was measured rectally with a

thermocouple probe (Physitemp). A 50% mechanical threshold was determined with von Frey hair filaments using the 'up and down paradigm' starting with 0.1 g and ending with a 2.0 g filament as a cut-off value. Noxious thermal thresholds were determined using an adaptation of the radiant heat test³², and cold/hot plate tests in which the temperature of the platform was varied between +52 to –10 °C. Time to first response for shivering, flinching or hind paw lifting was recorded with a cut-off value of 300 s for cold or 20 s for heat to prevent thermal injury. Responses to evaporative cooling were assessed by applying acetone to the hind paw as previously described¹¹. For the two-plate choice test, animals were placed in a chamber containing identical adjacent platforms with one set to 30 °C and the other adjusted to various temperatures. Mice were free to explore and the total time spent on each surface was recorded over a 5 min period. All data sets were analysed using two- or one-way ANOVA analysis followed by Tukey's HSD post-hoc analysis. All behavioural tests were performed by a single individual who was blind to genotype. Behavioural experiments were performed with F2 generation littermates and thus it is formally possible that strain-polymorphic determinants could contribute to cold preference behaviour if they are closely linked to the TRPM8 locus. Future analysis of animals on a more uniform genetic background will address this issue.

31. Bunting, M., Bernstein, K. E., Greer, J. M., Capecchi, M. R. & Thomas, K. R. Targeting genes for self-excision in the germ line. *Genes Dev.* 13, 1524–1528 (1999).
32. Hargreaves, K., Dubner, R., Brown, F., Flores, C. & Joris, J. A new and sensitive method for measuring thermal nociception in cutaneous hyperalgesia. *Pain* 32, 77–88 (1988).

An efflux transporter of silicon in rice

Jian Feng Ma¹, Naoki Yamaji^{1*}, Namiki Mitani^{1*}, Kazunori Tamai¹, Saeko Konishi², Toru Fujiwara^{3,4}, Maki Katsuhara¹ & Masahiro Yano⁵

Silicon is an important nutrient for the optimal growth and sustainable production of rice^{1–4}. Rice accumulates up to 10% silicon in the shoot, and this high accumulation is required to protect the plant from multiple abiotic and biotic stresses^{1–5}. A gene, *Lsi1*, that encodes a silicon influx transporter has been identified in rice⁶. Here we describe a previously uncharacterized gene, *low silicon rice 2* (*Lsi2*), which has no similarity to *Lsi1*. This gene is constitutively expressed in the roots. The protein encoded by this gene is localized, like *Lsi1*, on the plasma membrane of cells in both the exodermis and the endodermis, but in contrast to *Lsi1*, which is localized on the distal side, *Lsi2* is localized on the proximal side of the same cells. Expression of *Lsi2* in *Xenopus* oocytes did not result in influx transport activity for silicon, but preloading of the oocytes with silicon resulted in a release of silicon, indicating that *Lsi2* is a silicon efflux transporter. The identification of this silicon transporter revealed a unique mechanism of nutrient transport in plants: having an influx transporter on one side and an efflux transporter on the other side of the cell to permit the effective transcellular transport of the nutrients.

Silicon is the second most abundant element in the Earth's crust and soil. It is essential for diatoms and plants in the Equisetales and is 'quasi-essential' for higher-plant growth³. All plants contain silicon, but the concentration differs greatly with the plant species, ranging from 0.1 to 10% in shoots^{1,7}. Rice (*Oryza sativa* L.) is a typical silicon-accumulating plant^{5,7}, requiring large amounts of silicon for vigorous growth and high production^{4,7}. Silicon deposited in the leaves, stem and husk forms a physiological barrier⁸, which mitigates fungal infection and pest attack, alleviates lodging and other abiotic stress, improves the light-interception ability by plants in a community,

and minimizes transpiration losses⁷. Soluble silicon in plants also has an active function in enhancing host resistance to plant diseases by stimulating one or more defence reaction mechanisms⁹.

Rice takes up much more silicon from the roots than other graminaceous crops, and the involvement of a specific transporter for silicon has been proposed^{10,11}. Recently, a gene, *Lsi1*, encoding a silicon uptake transporter was identified in rice⁶. This gene was cloned by using a rice mutant defective in silicon uptake¹². The protein belongs to a Nod26-like major intrinsic protein (NIP) subfamily in the aquaporin family. *Lsi1* is expressed mainly in the roots, and the encoded protein shows transport activity specific to silicon. The transport protein is localized on the plasma membrane of the distal side of cells in both the exodermis and endodermis. *Lsi1* is an influx transporter responsible for the transport of silicon from the external solution to the root cells. Therefore, to reach the stele, silicon must be transported out of the cells. However, genes responsible for the efflux of silicon have not been identified until now.

To identify genes involved in the efflux transport of silicon, we screened new rice mutants defective in silicon uptake, using tolerance to germanium as a selection parameter. Germanium is an analogue of silicon, and plant roots take up germanium in a manner similar to that of silicon^{13–15}. However, germanium is toxic to plants, producing brown spots on the leaves¹². We obtained a mutant that had no brown spots on the leaves from about 4,600 M₃ seeds (cv. Taichung-65) mutagenized with *N*-methyl-*N*-nitrosourea and named this mutant *low silicon rice 2* (*Lsi2*). The uptake of silicon by the mutant was much lower than that by wild-type rice (Fig. 1a). The morphology and growth of this mutant at a young stage were the same as those of wild-type rice when they were grown in the absence of silicon. However, when they were grown in the field, both growth and

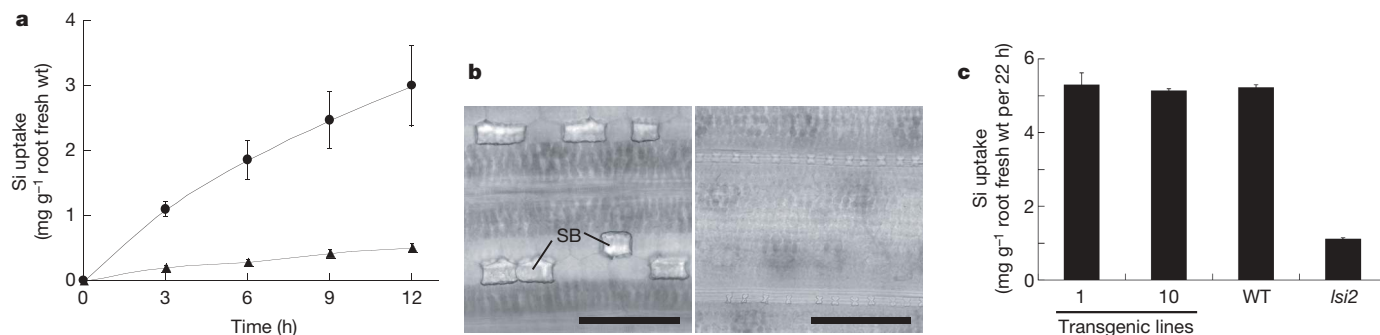


Figure 1 | Silicon uptake and phenotype of *Lsi2* mutant. **a**, Time-dependent uptake of silicon in wild-type rice (circles; cv. Taichung-65) and the *Lsi2* mutant (triangles). Data are shown as means \pm s.d. ($n = 3$). **b**, Silica bodies (SB) on the flag leaf of wild-type rice (left) and the *Lsi2* mutant (right) grown

in a field. Scale bars, 100 μ m. **c**, Complementation test showing silicon uptake by two independent transgenic lines as well as the wild-type rice (WT) and the mutant. Data are shown as means and s.d. ($n = 3$).

¹Research Institute for Bioresources, Okayama University, Chuo 2-20-1, Kurashiki 710-0046, Japan. ²Institute of Society for Techno-innovation of Agriculture, Forestry and Fisheries, Kamiyokoba, Tsukuba, Ibaraki, 305-0854, Japan. ³Biotechnology Research Center, The University of Tokyo, Bunkyo Ku, Tokyo, 113-8657, Japan. ⁴Solution Oriented Research for Science and Technology, Japan Science and Technology Agency, 3-4-15 Nihonbashi, Tokyo 103-0027, Japan. ⁵QTL Genomics Research Center, National Institute of Agrobiological Sciences, Tsukuba, Ibaraki 305-8602, Japan.

*These authors contributed equally to this work.

production were repressed in the mutant (Supplementary Fig. 1); the grain yield of the mutant was only 40% of that of the wild-type rice. The decrease resulted from low accumulation of silicon in the shoot and hull in the mutant (Fig. 1b, and Supplementary Fig. 2). There was no difference in uptake of potassium and phosphorus between the mutant and the wild-type rice (Supplementary Fig. 2). Genetic analysis showed that the low uptake of silicon in the mutant was controlled independently of *Lsi1* by a recessive gene (Supplementary Fig. 3).

We constructed an F₂ mapping population derived from a cross between *Lsi2* and Kasalath, an *indica* cultivar. We mapped the gene (*Lsi2*) roughly to the distal region of chromosome 3 with bulk segregant analysis by using F₂ plants (Supplementary Fig. 4a). For fine mapping of *Lsi2* we used about 300 homozygotes with a low uptake of silicon. We developed new markers and mapped the candidate region of *Lsi2* to 106 kilobases (kb) between the markers *lsi2-b* and *lsi2-c* (Supplementary Fig. 4b). There are predicted to be 17 genes in this region. On the basis of the annotation, we chose a gene annotated with a putative anion transporter as a candidate (Supplementary Fig. 4c) and then sequenced and compared the gene in the wild type with that in the *Lsi2* mutant. We found a single-base mutation in the first exon of the candidate gene from G in the wild type to A in the mutant, resulting in an amino acid change from serine 115 to asparagine (Supplementary Fig. 4d, 5). The gene consists of two exons and one intron (Supplementary Fig. 4d). The open reading frame (ORF) of this gene was 1,416 base pairs long and the deduced protein consisted of 472 amino-acid residues (Supplementary Fig. 5). The gene is predicted to encode a membrane protein with 11 transmembrane domains.

To confirm that this candidate gene was *Lsi2*, we performed a complementation test by introducing an 8-kilobase-pair clone containing the promoter and the candidate gene into the *Lsi2* mutant. In the two independently generated T₀ transgenic lines, silicon uptake was completely recovered, but not in the control lines. Furthermore, in T₁ lines carrying *Lsi2*, silicon uptake was the same as that in the wild type (Fig. 1c). These results confirm that the candidate gene is *Lsi2* and that it is responsible for silicon uptake.

A BLAST search and a ClustalW analysis revealed that *Lsi2* is a putative anion transporter with no similarity to the silicon influx transporter *Lsi1* (ref. 6). In all we found six full-length homologues in higher plants in the database, including one in *Arabidopsis* and five in rice (Supplementary Fig. 6). In addition, some close homologues were reported as expressed sequence tag clones from other graminaceous crops, including wheat, barley, maize, sorghum and sugar cane. The closest homologue in rice (Os10g0547500) has an identity of 81% with *Lsi2* (Supplementary Fig. 6). The functions of these genes, however, have not yet been identified in plants. In *Escherichia coli*, a gene (*ArsB*) with the same motif as *Lsi2* encodes an efflux transporter of arsenic and antimony¹⁶. Whether *Lsi2* has transport activity for arsenic remains to be determined, although rice accumulates silicon much more than arsenic.

Lsi2 was expressed mainly in the roots (Fig. 2a). There was no difference in the mRNA accumulation pattern between *Lsi2* (the wild type) and *lsi2* (the mutant). The mRNA accumulation was constitutive but it was decreased by one-quarter by a continuous supply of silicon for three days (Supplementary Fig. 7). Furthermore, the accumulation of *Lsi2* transcripts was very low in the root tip (0–10 mm), but high in the mature parts of the roots (Supplementary Fig. 8). Because the transcripts of *Lsi1* showed a similar accumulation pattern⁶, the expression of *Lsi1* and *Lsi2* may be regulated in a similar manner. In fact, we found that there are common domains in the promoter region between *Lsi1* and *Lsi2* (Supplementary Fig. 9).

We investigated the subcellular localization of *Lsi2* by delivering a translational fusion between green fluorescent protein (GFP) and *Lsi2* and into onion epidermal cells by particle bombardment. Cells expressing the GFP–*Lsi2* fusion showed a GFP signal only at the

plasma membrane, which is located inside the cell wall but outside the nucleus and vacuole (Supplementary Fig. 10).

For further examination of the localization of *Lsi2* in rice roots, we made transgenic rice plants carrying GFP under the control of the *Lsi2* promoter region (2 kb). The GFP fluorescence signal was observed in both the exodermis and the endodermis (Fig. 2b). Furthermore, we stained the roots with an anti-*Lsi2* polyclonal antibody as well as co-staining with an anti-*Lsi1* polyclonal antibody and 4,6-diamidino-2-phenylindole (DAPI). We also observed that *Lsi2* was localized at the plasma membrane of both the exodermal and endodermal cells of the roots (Fig. 2c, d). *Lsi2* is therefore localized at the same cells as *Lsi1* (Fig. 2d, ref. 6). However, in contrast to *Lsi1*, which is localized on the distal side⁶, *Lsi2* was localized on the proximal side of the exodermal and endodermal cells (Fig. 2c, d).

To examine the silicon transport activity of *Lsi2*, we injected water or capped RNA encoding *Lsi1* or *Lsi2* into *Xenopus laevis* oocytes and then measured the silicon inside the oocytes after incubation for 2 h in a solution containing silicic acid labelled with ⁶⁸Ge. In contrast to *Lsi1*, which showed uptake activity into oocytes, *Lsi2* did not show uptake activity for silicon (Supplementary Fig. 11), suggesting that *Lsi2* is unable to transport silicon from the external solution into the cell. However, when silicic acid labelled with ⁶⁸Ge was preloaded into the oocytes, release of silicon was significantly increased in the oocytes injected with the wild-type *Lsi2* (Fig. 3a). These results suggest that in contrast to influx transporter *Lsi1* (designated SIIT1), *Lsi2* is a silicon efflux transporter (SIET1), which is capable of transporting silicon out of the cells. In oocytes injected with mutant *lsi2*, no efflux of silicon occurred (Fig. 3a). Computer modelling predicted that the residue change at the 115 position in the mutant may affect the conformation change of the protein that is required when silicon is transported. Efflux of silicon was inhibited by low-temperature treatment and by three protonophores, namely 2,4-dinitrophenol (DNP), carbonylcyanide 3-chlorophenylhydrazone (CCCP) and carbonylcyanide *p*-(trifluoromethoxy)phenylhydrazone (FCCP) (Supplementary Fig. 12). This is in agreement with the findings obtained in rice¹². Furthermore, the efflux increased with decreasing external pH values (Fig. 3b). Taken together, these results suggest that the transport of silicon by SIET1 is an energy-dependent

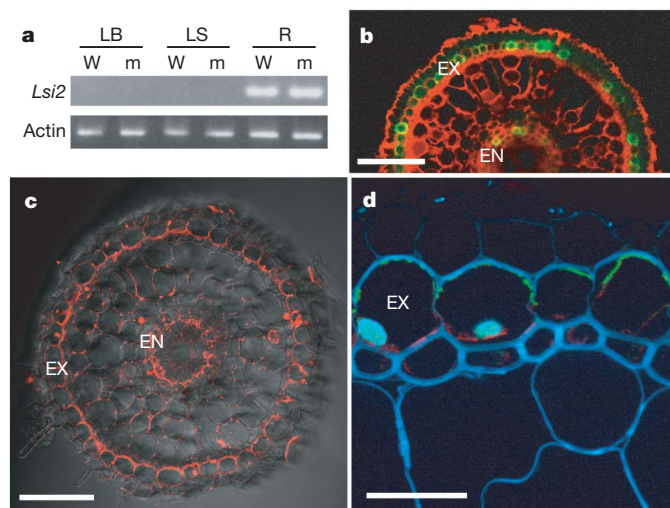


Figure 2 | Expression and localization of *Lsi2*. **a**, Analysis of *Lsi2* and actin-gene (internal standard) transcripts in the leaf blade (LB), leaf sheath (LS) and root (R) in the wild-type (W) and *Lsi2* mutant (m) by RT-PCR. **b**, Fluorescence of GFP controlled by the *Lsi2* promoter in transgenic plants. Cross-section observed by laser scanning confocal microscopy, counterstained with propidium iodide. **c**, Immunolocalization of *Lsi2* in the seminal root (red). **d**, Merged image of *Lsi1* (green), *Lsi2* (red) immunolocalization, nuclei stained by DAPI (cyan) and cell wall autofluorescence (blue). EX, exodermis; EN, endodermis. Scale bars, 100 μm (b, c), 20 μm (d).

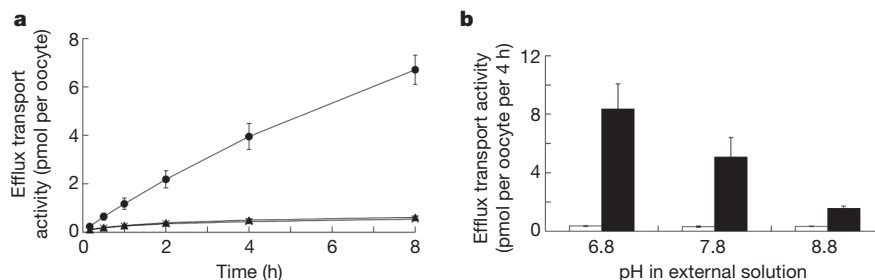


Figure 3 | Efflux activity of Lsi2 in *Xenopus* oocytes. **a**, Time-dependent efflux of silicon. Oocytes previously injected with water (crosses) or with Lsi2 (circles) or Lsi2 (triangles) cRNA were preloaded with 1.0 mM silicon labelled

with ^{68}Ge . Data are shown as means \pm s.d. ($n = 3$ or 4). **b**, Dependence of silicon efflux on extracellular pH. Open bars, water; filled bars, Lsi2. Data are shown as means and s.d. ($n = 3$ or 4).

active process driven by the proton gradient, although more work will be required to determine the actual transport mechanisms unambiguously.

Casparian strips exist in both the exodermis and endodermis in rice⁶. Because of the casparian strips, solutes cannot pass freely from the external solution to the stele¹⁷. In fact, large amounts of silicon are deposited on the exodermis and endodermis¹⁸. Both SIIT1 and SIET1 are required for the transcellular transport of silicon to the stele (Supplementary Fig. 13). Moreover, in rice roots aerenchyma is formed¹⁹, accompanied by the destruction of almost all cortex cells except the exodermis and endodermis in mature roots. Therefore, silicon transported into the exodermal cells by SIIT1 is released by SIET1 into the apoplast of a spoke-like structure across the aerenchyma (Supplementary Fig. 13). Silicon is then transported into the endodermal cells by SIIT1 and released into the stele by SIET1 (Supplementary Fig. 13). Coupling of SIIT1 and SIET1 in the same cell in the casparian strips is therefore required for the efficient transport of silicon across the cells into the stele.

In the plant kingdom, silicon is accumulated highly in the Bryophyta, as well as the Lycopodiophyta and Equisetophyta, but is accumulated hardly at all in the Filicopsida in the Pteridophyta, and in the Gymnospermae²⁰. In the Angiospermae, only the Cyperaceae and Gramineae show a high accumulation of silicon^{20,21}. However, only one gene (*Lsi1*) responsible for silicon uptake has so far been identified in higher plants⁶. In the marine diatom *Cylindrotheca fusiformis*, which requires silicon as an essential element, a gene family encoding a silicon transporter has been identified^{22,23}. However, there is no similarity between these genes and *Lsi2* identified here. The identification of *Lsi2* as well as *Lsi1* not only provides a new insight into the silicon uptake system in higher plants; it also provides a clue to the evolutionary relationship of silicon accumulation.

METHODS SUMMARY

We obtained a rice mutant (*Lsi2*) defective in silicon uptake by screening mutants from M_3 seeds treated with the mutagen *N*-methyl-*N*-nitrosourea (wild-type Taichung-65) on the soil containing 50 mg of Ge per kg of soil, as GeO_2 . We determined silicon uptake by the *Lsi2* mutant and the wild-type rice by subjecting young seedlings grown hydroponically to a nutrient solution containing 0.5 mM silicon as silicic acid, as described previously²⁴. We also grew both the wild-type rice and the *Lsi2* mutant in a field from June to October. We investigated the grain yield and observed the silica bodies of the flag leaf by using safranin staining²⁴. The concentrations of minerals (silicon, phosphorus and potassium) in the shoots and hulls were determined as reported²⁵.

We first roughly mapped *Lsi2* by bulk segregant analysis with F_2 plants derived from a cross between *Lsi2* and Kasalath, as described previously⁶. For fine mapping we used a total of 300 homozygote plants showing low uptake of silicon. For the complementation test, we transferred an 8-kb clone containing *Lsi2* and its promoter into the calluses derived from the mutant *Lsi2*, using an *Agrobacterium*-mediated transformation system²⁶. To investigate the cellular localization of *Lsi2* we introduced a construct consisting of the promoter (2 kb) of *Lsi2* fused with *GFP* to calluses (cv. Nipponbare).

We also examined the cellular localization of *Lsi2* with immunohistological fluorescence staining as described previously⁶. For multi-colour staining of *Lsi1* and *Lsi2*, we used a Zenon Rabbit IgG Labelling Kit (Molecular Probes) in accordance with the manufacturer's instructions.

We investigated the expression of *Lsi2* in different tissues by quantitative RT-PCR. We measured the transport activity of silicon by exposing *Xenopus* oocytes injected with cRNA of *Lsi1*, *Lsi2* or *Lsi2* (50 nl, 1 ng nl⁻¹) or distilled water to ^{68}Ge -labelled silicon solution under different conditions. We then measured the radioactivity in the external solution and the oocytes with a liquid scintillation counter²⁷.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 6 November 2006; accepted 31 May 2007.

- Epstein, E. Silicon. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **50**, 641–664 (1999).
- Ma, J. F. Role of silicon in enhancing the resistance of plants to biotic and abiotic stresses. *Soil Sci. Plant Nutr.* **50**, 11–18 (2004).
- Epstein, E. & Bloom, A. J. in *Mineral Nutrition of Plants: Principles and Perspectives* 227–231 (Sinauer Associates, Sunderland, Massachusetts, 2005).
- Savant, N. K., Snyder, G. H. & Datnoff, L. E. Silicon management and sustainable rice production. *Adv. Agron.* **58**, 151–199 (1997).
- Ma, J. F. & Yamaji, N. Silicon uptake and accumulation in higher plants. *Trends Plant Sci.* **11**, 392–397 (2006).
- Ma, J. F. *et al.* A silicon transporter in rice. *Nature* **440**, 688–691 (2006).
- Ma, J. F. & Takahashi, E. in *Soil, Fertilizer, and Plant Silicon Research in Japan* 107–180 (Elsevier Science, Amsterdam, 2002).
- Yoshida, S. Chemical aspects of the role of silicon in physiology of the rice plant. *Bull. Natl Inst. Agric. Sci. B* **15**, 1–58 (1965).
- Fauteux, F., Remus-Borel, W., Menzies, J. G. & Belanger, R. R. Silicon and plant disease resistance against pathogenic fungi. *FEMS Microbiol. Lett.* **249**, 1–6 (2005).
- Tamai, K. & Ma, J. F. Characterization of silicon uptake by rice roots. *New Phytol.* **158**, 431–436 (2003).
- Raven, J. A. in *Silicon in Agriculture* (eds Datnoff, L. E., Snyder, G. H. & Korndörfer, G. H.) 41–55 (Elsevier Science, Amsterdam, 2001).
- Ma, J. F., Tamai, K., Ichii, M. & Wu, K. A rice mutant defective in Si uptake. *Plant Physiol.* **130**, 2111–2117 (2002).
- Takahashi, E., Syo, S. & Miyake, Y. Effect of germanium on the growth of plants with special reference to the silicon nutrition (Part 1). *J. Sci. Soil Manure Jpn* **47**, 183–190 (1976).
- Takahashi, E., Syo, S. & Miyake, Y. Effect of germanium on the growth of plants with special reference to the silicon nutrition (Part 2). *J. Sci. Soil Manure Jpn* **47**, 191–197 (1976).
- Rains, D. W., Epstein, E., Zasoski, R. J. & Aslam, M. Active silicon uptake by wheat. *Plant Soil* **280**, 223–228 (2006).
- Meng, Y. L., Liu, Z. & Rosen, B. P. As(III) and Sb(III) uptake by GlpF and efflux by ArsB in *Escherichia coli*. *J. Biol. Chem.* **279**, 18334–18341 (2004).
- Taiz, L. & Zeiger, E. in *Plant Physiology* 2nd edn, 103–124 (Sinauer Associates, Sunderland, 1998).
- Gong, H. J., Randall, D. P. & Flowers, T. J. Silicon deposition in the root reduces sodium uptake in rice (*Oryza sativa* L.) seedlings by reducing bypass flow. *Plant Cell Environ.* **29**, 1970–1979 (2006).
- Kawai, M., Samarajeewa, P. K., Barrero, R. A., Nishiguchi, M. & Uchimiya, H. Cellular dissection of the degradation pattern of cortical cell death during aerenchyma formation of rice roots. *Planta* **204**, 277–287 (1998).
- Takahashi, E., Ma, J. F. & Miyake, Y. The possibility of silicon as an essential element for higher plants. *Comments Agric. Fd Chem.* **2**, 99–122 (1990).
- Hodson, M. J., White, P. J., Mead, A. & Broadley, M. R. Phylogenetic variation in the silicon composition of plants. *Ann. Bot.* **96**, 1027–1046 (2005).

22. Hildebrand, M., Higgins, D. R., Busser, K. & Volcani, B. E. Silicon-responsive cDNA clones isolated from the marine diatom *Cylindrotheca fusiformis*. *Gene* **132**, 213–218 (1993).
23. Hildebrand, M., Volcani, B. E., Gassmann, W. & Schroeder, J. I. A gene family of silicon transporters. *Nature* **385**, 688–689 (1997).
24. Ma, J. F., Goto, S., Tamai, K. & Ichii, M. Role of root hairs and lateral roots in silicon uptake by rice. *Plant Physiol.* **127**, 1773–1780 (2001).
25. Ma, J. F., Higashitani, A., Sato, K. & Tateda, K. Genotypic variation in Si content of barley grain. *Plant Soil* **249**, 383–387 (2003).
26. Hiei, Y., Ohta, S., Komari, T. & Kumashiro, T. Efficient transformation of rice (*Oryza sativa* L.) mediated by *Agrobacterium* and sequence analysis of the boundaries of the T-DNA. *Plant J.* **6**, 271–282 (1994).
27. Tallberg, P., Koski-Vahala, J. & Hartikainen, H. Germanium-68 as a tracer for silicon fluxes in freshwater sediment. *Water Res.* **36**, 956–962 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank H. Sato and Y. Nagato for providing M₃ seeds, and M. Ishiguro for computer modelling. This research was supported by a Grant-in-Aid for Scientific Research on Priority Areas from the Ministry of Education, Culture, Sports, Science and Technology of Japan (to J.F.M.) and a grant of Green Technology Project from the Ministry of Agriculture, Forestry and Fisheries of Japan (to J.F.M.).

Author Contributions N.M. and N.Y. contributed equally to this work. J.F.M. and K.T. cloned the gene *Lsi2* with the help of S.K. and M.Y.; N.Y. investigated the localization of *Lsi2*; and N.M. measured the transport activity of *Lsi2*. J.F.M. performed screening and physiological experiments and wrote the paper. All authors discussed the results and commented on the manuscript.

Author Information The nucleotide sequence data reported in this paper are deposited in the DDBJ/EMBL/GenBank nucleotide sequence databases under the accession number AB222273. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.F.M. (maj@rib.okayama-u.ac.jp).

METHODS

Screening of rice mutants defective in silicon uptake. We sowed M_3 seeds treated with the mutagen *N*-methyl-*N*-nitrosourea (wild-type Taichung-65) on soil containing 50 mg of Ge per kg of soil, as GeO_2 . After a one-month cultivation, we selected seedlings that did not show brown spots on the leaf blades¹². We then grew the candidate mutants in the field and obtained a novel mutant named *low silicon rice 2* (*lsi2*) after secondary screening by determination of silicon uptake.

Physiological study of *lsi2*. We determined silicon uptake by the *lsi2* mutant and by the wild type by subjecting young seedlings grown hydroponically in a nutrient solution containing 0.5 mM silicon as silicic acid, as described previously²⁴. We also grew both wild-type rice and the *lsi2* mutant in a field from June to October. We investigated the grain yield and took the flag leaf for the detection of silica bodies at harvest. Silica bodies were observed under a microscope after staining of the leaf with 0.01% safranin in boiled phenol²⁴. The concentrations of minerals (silicon, phosphorus and potassium) in the shoots and hulls were determined as reported²⁵.

Map-based cloning of *lsi2*. We first mapped *lsi2* roughly by bulk segregant analysis with F_2 plants derived from a cross between *lsi2* and Kasalath, as described previously⁶. For fine mapping we used a total of 300 homozygote plants showing low uptake of silicon. Linkage analysis with both simple sequence repeats (SSR) and PCR markers revealed that *lsi2* was restricted to a 106-kb genomic region on the PAC clone AC113930.

Complementation test. We obtained a clone (K0135E16) containing *lsi2* by screening a total of 48,000 bacterial artificial chromosome clones of Kasalath. After partial enzymatic cutting with *Sau3AI*, we selected an 8-kb clone containing *lsi2* and its promoter. This clone was inserted into a vector (pPZP2H-lac)²⁸, and then transformed into *Agrobacterium tumefaciens* (strain EHA101). We transferred it into calluses derived from the mutant *lsi2* by using an *Agrobacterium*-mediated transformation system²⁶. We obtained two independent transgenic lines and vector control lines each and determined silicon uptake in the T_0 and T_1 lines in a nutrient solution containing 1.0 mM silicon.

Cellular localization of *lsi2*. We first investigated the subcellular localization of *lsi2* by expressing it transiently in onion epidermal cells. Onion epidermal cells were bombarded with 1- μ m gold particles coated with plasmid DNA carrying *lsi2* fused with GFP incubated in the dark at 25 °C for 20 h. After plasmolysis of cells with 1 M mannitol, we observed fluorescence by laser-scanning confocal microscopy.

To investigate the cellular localization of *lsi2* further, we introduced a construct consisting of the promoter (2 kb) of *lsi2* fused with *GFP* to calluses (cv. Nipponbare) by using an *Agrobacterium*-mediated transformation system²⁶. We selected transformed calluses by hygromycin resistance, and from them regenerated plants.

We also examined the cellular localization of *lsi2* by immunohistological fluorescence staining, as described previously⁶. We observed fluorescence with a laser scanning confocal microscope (LSM510; Carl Zeiss). For multi-colour staining of *lsi1* and *lsi2*, we used a Zenon Rabbit IgG Labelling Kit (Molecular Probes) in accordance with the manufacturer's instructions. We also stained the root sections with DAPI and observed the autofluorescence of the cell wall.

Expression studies. We investigated the expression of *lsi2* by quantitative RT-PCR. We extracted total RNA from the leaf blade, leaf sheath and root as well as different segments of the root, and then converted them to complementary DNA. *lsi2*, and the actin gene as an internal control, were amplified by using SYBR Green I real-time PCR with the following primer pairs: 5'-ATCTGG-GACTTCATGGCCC-3' (forward) and 5'-ACGTTTGATGCGAGGTTGG-3' (reverse) for *lsi2*, and 5'-GACTCTGGTGATGGTGTTCAGC-3' (forward) and 5'-GGCTGGAAGAGGACCTCAGG-3' (reverse) for the actin gene.

Transport activity in oocytes. We performed *in vitro* transcription from a pXβG-ev1 poly(A) vector carrying *lsi1*, *lsi2* or *lsi2* cDNA. We injected distilled water or the capped cRNA of *lsi1*, *lsi2* or *lsi2* (50 nl, 1 ng nl⁻¹) into *Xenopus* oocytes selected according to size and development stage. For uptake experiments, after incubation for one day, the oocytes were exposed to a solution containing 1 mM silicon labelled with ⁶⁸Ge (9 MBq mmol⁻¹) for 2 h (ref. 27). We then washed the oocytes five times with modified Barth's saline (MBS) without silicon and ⁶⁸Ge, and homogenized them with 0.1 M HNO₃. For efflux experiments we injected 50 nl of solution containing 1 mM silicon labelled with ⁶⁸Ge (2 GBq mmol⁻¹) into oocytes. We washed the oocytes for 10 min in MBS without silicon and ⁶⁸Ge to remove extracellular silicon and ⁶⁸Ge, and then transferred them to a new tube containing 500 μ l of fresh MBS at 10 or 18 °C in the presence or absence of inhibitors, namely 0.5 mM DNP, 10 μ M CCCP or 10 μ M FCCP, or at different pH values. After 4 h the incubation solution was carefully sampled. Finally, after one day we measured the radioactivity in the external solution and oocytes with a liquid scintillation counter.

28. Fuse, T., Sasaki, T. & Yano, M. Ti-plasmid vectors useful for functional analysis of rice genes. *Plant Biotechnol.* **18**, 219–222 (2001).

Antitumour drugs impede DNA uncoiling by topoisomerase I

Daniel A. Koster¹, Komaraiah Palle², Elisa S. M. Bot¹, Mary-Ann Bjornsti² & Nynke H. Dekker¹

Increasing the ability of chemotherapeutic drugs to kill cancer cells is often hampered by a limited understanding of their mechanism of action. Camptothecins, such as topotecan, induce cell death by poisoning DNA topoisomerase I, an enzyme capable of removing DNA supercoils^{1–4}. Topotecan is thought to stabilize a covalent topoisomerase–DNA complex^{5–7}, rendering it an obstacle to DNA replication forks^{2,3,8,9}. Here we use single-molecule nanomanipulation to monitor the dynamics of human topoisomerase I in the presence of topotecan. This allowed us to detect the binding and unbinding of an individual topotecan molecule in real time and to quantify the drug-induced trapping of topoisomerase on DNA. Unexpectedly, our findings also show that topotecan significantly hinders topoisomerase-mediated DNA uncoiling, with a more pronounced effect on the removal of positive (overwound) versus negative supercoils. *In vivo* experiments in the budding yeast verified the resulting prediction that positive supercoils would accumulate during transcription and replication as a consequence of camptothecin poisoning of topoisomerase I. Positive supercoils, however, were not induced by drug treatment of cells expressing a catalytically active, camptothecin-resistant topoisomerase I mutant. This combination of single-molecule and *in vivo* data suggests a cytotoxic mechanism for camptothecins, in which the accumulation of positive supercoils ahead of the replication machinery induces potentially lethal DNA lesions.

DNA topoisomerases resolve topological problems by means of transient DNA strand breakage and religation^{1,4,10}. However, drug-stabilized topoisomerase–DNA complexes may also induce potentially lethal DNA damage. Indeed, eukaryotic DNA topoisomerase IB (TopIB) is the cellular target of the camptothecin class of chemotherapeutics^{2,3,11}. The camptothecin analogues topotecan and irinotecan have significant activity against adult and paediatric solid tumours, and have gained US Food and Drug Administration approval for the treatment of ovarian and small-cell lung cancer^{12–14}. A detailed understanding of topoisomerase–drug interactions is critical for optimal clinical development of these chemotherapeutics. However, the dynamic interactions underlying this poisoning and their biological ramifications remain largely unknown.

TopIB removes DNA supercoils by first forming a clamp around duplex DNA^{15,16}. The active-site tyrosine acts as a nucleophile to cleave a single DNA strand, forming a transient DNA–(3′-phosphotyrosyl)-enzyme ‘covalent complex’ and a free 5′-OH DNA end. Torsional energy within the DNA then drives uncoiling about the intact DNA strand. Following the removal of a random number of supercoils, a ligation reaction restores the DNA backbone^{17,18}. Topotecan intercalates into the nick generated by TopIB, thereby preventing religation and trapping TopIB on the DNA^{5–7,19}. During S phase, these reversible ternary topotecan–TopIB–DNA complexes are converted into cytotoxic DNA lesions that cause cell death^{2,3,8,9}. Because DNA replication is required for topotecan-induced cell lethality^{8,9}, it has been proposed

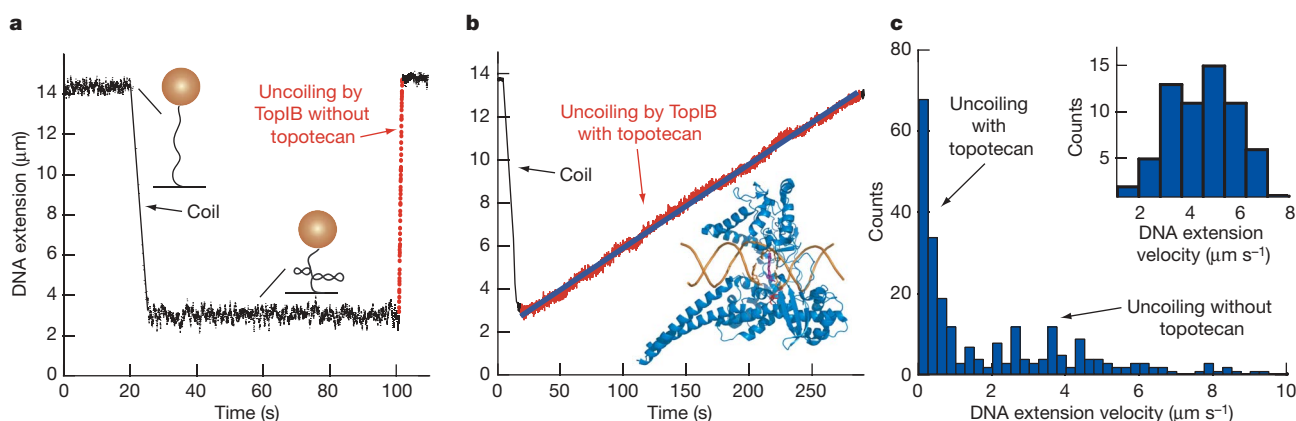


Figure 1 | Drug-dependent slow DNA uncoiling by human TopIB. **a**, In the absence of topotecan, uncoiling (red dots, at 100 s) proceeds rapidly. **b**, In the presence of topotecan, slow and continuous uncoiling of positive supercoils is observed. The uncoiling rate is constant, as expected in the constant torque regime (Supplementary Information IV). Inset, ribbon diagram of the carboxy terminal 70 kDa of human TopIB (blue) covalently linked to DNA (yellow) in the presence of topotecan (magenta). **c**, The

distribution of uncoiling velocities in the presence of 5 μM topotecan ($n = 275$) and $F = 0.2$ pN shows contributions from topotecan-mediated slow uncoiling and topotecan-independent uncoiling (inset, measurement in the absence of topotecan ($n = 64$)). Topotecan-mediated uncoiling of positive supercoils proceeds ~20-fold more slowly than topotecan-independent uncoiling.

¹Kavli Institute of Nanoscience, Faculty of Applied Sciences, Delft University of Technology, Lorentzweg 1, 2628 CJ Delft, The Netherlands. ²Department of Molecular Pharmacology, St Jude Children's Research Hospital, 332 N. Lauderdale, Memphis, Tennessee 38105, USA.

that these lesions occur as a result of the covalent complex 'colliding' with an advancing replication fork^{2,3}. Here we describe single-molecule experiments that test key predictions of the proposed mechanism of poisoning of TopIB, and reveal dynamic interactions between topotecan and TopIB. The consequences of these interactions are also observed in living cells, suggesting an alternative mechanism for cell death.

To examine the consequences of camptothecin binding on TopIB swivel function, we explored its real-time dynamics in the presence of topotecan in the context of a single molecule. Our experimental configuration, the magnetic tweezers, is described elsewhere²⁰ (Supplementary Information I). In a typical experiment the DNA is mechanically coiled in the presence of TopIB, introducing multiple plectonemes that reduce the DNA extension (Fig. 1a). We usually observe a plateau (for example, from ~25 to 100 s in Fig. 1a), which is indicative of DNA in a supercoiled and unknicked state. In the presence of human TopIB (encoded by the *TOPI* gene) but the absence of topotecan, the DNA is subsequently (at 100 s, shown in red) rapidly uncoiled by the enzyme. Plectoneme removal occurs either in a single enzymatic event or in multiple steps (Supplementary Information II). Surprisingly, on addition of topotecan, a very different signature is observed (Fig. 1b). First, topotecan-mediated uncoiling occurs slowly (red points) compared with uncoiling in the absence of drug. Slow uncoiling is observed immediately following mechanical coiling; this proceeds in a fashion that seems to be continuous, and can be fitted by a linear relation (blue line). Second, slow uncoiling continues for long periods of time. Finally, during the slow uncoiling, no plateaux are apparent, suggesting that religation does not take place (Supplementary Information III). Structural studies revealed that the ring-like structure of topotecan intercalates into a TopIB-generated nick and is stabilized by hydrogen bonds to TopIB and base-stacking

interactions with adjacent bases^{7,19}. Consistent with biochemical assays of camptothecin-induced covalent TopIB–DNA complexes^{5,21}, structural studies also show that topotecan locally deforms the DNA duplex and displaces the 5'-OH, presumably decreasing the probability of religation. It was also speculated that contacts of topotecan with TopIB and DNA ought to significantly hinder rotational motion^{7,16,22}—a prediction borne out by our observations.

To quantify the reduction in uncoiling rate caused by topotecan binding, we examined the complete distribution of DNA extension velocities during enzymatic uncoiling (Fig. 1c). In the presence of topotecan, we identified two populations: one topotecan-mediated and one non-topotecan-mediated. The fraction of topotecan-mediated events increases with drug concentration (data not shown), but, even at the highest concentrations, a significant fraction of events remains non-topotecan-mediated. These findings are consistent with biochemical studies demonstrating that topotecan stabilizes a subset of TopIB–DNA complexes^{2,3}. Non-topotecan-mediated events, unambiguously identified by experiments in the absence of topotecan (Fig. 1c, inset), take place, on average, at $4.1 \mu\text{m s}^{-1}$. Topotecan-mediated events, however, take place, on average, at $0.2 \mu\text{m s}^{-1}$ (corresponding to a plectoneme relaxation rate of ~3 Hz; Supplementary Information IV), independent of drug concentration. We conclude that topotecan-mediated uncoiling by TopIB occurs roughly 20-fold slower than uncoiling by TopIB alone.

Control experiments were performed to exclude the possibility that topotecan binding to DNA, or non-specific TopIB–topotecan interactions could result in such slow uncoiling. First, topotecan binding did not affect the mechanical properties of DNA (Supplementary Information V). Second, slow uncoiling was not specific to topotecan, occurring also with camptothecin. However, it did not occur with Top1(Y723F), a catalytically inactive human TopIB

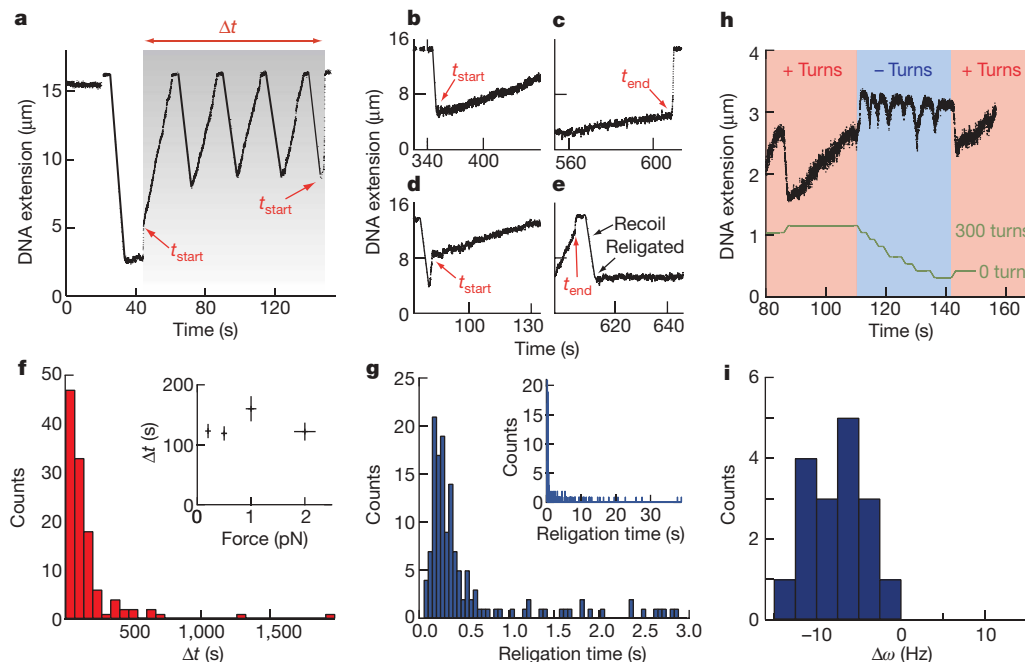


Figure 2 | The duration of slow removal and its uniqueness to the removal of positive supercoils. **a**, The duration Δt for a bound topotecan (grey area) is defined as the time period between t_{start} and t_{end} during which slow and constant uncoiling is observed. **b**, **c**, Typical illustration (70% of events, $n = 660$) of t_{start} (**b**) and t_{end} (**c**) events. **d**, t_{start} event in which nicking appears directly before the onset of slow uncoiling. **e**, Event suggestive of a topotecan molecule exiting at t_{end} , after which religation occurs (plateau onset at 610 s). **f**, Histogram of Δt for $F = 0.5$ pN, with mean of 121 ± 11 s ($n = 146$), which is not significantly force-dependent (inset). Means are numerical averages with the corresponding s.e.m. **g**, Distribution of TopIB religation times, with a most probable religation time of 0.3 s ($n = 132$) and a

slowly decaying tail (inset), giving an average time of 4 s, far below Δt (121 s, as shown in **f**). **h**, The removal rate of positive (pink areas) and negative (blue areas) supercoils in the presence of topotecan. Removal of positive supercoils proceeds more slowly. Following the removal of negative supercoils, continued slow removal of positive supercoils is verified to ascertain that the topotecan molecule remained bound in the interim. The rotational position of the magnets, specifying the absolute number of turns that are mechanically injected into the DNA, is indicated in green. **i**, Histogram of the difference in uncoiling velocity, $\Delta\omega$, between positive and negative supercoils, in the presence of topotecan.

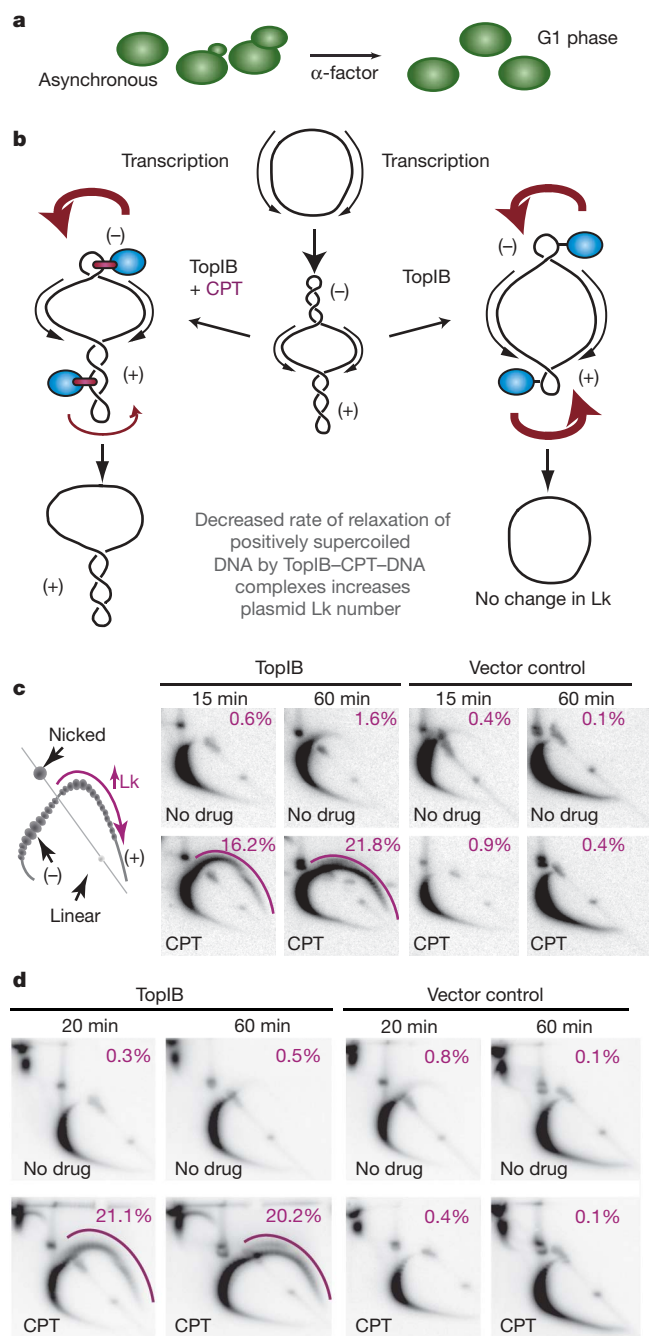


Figure 3 | Camptothecin poisoning of TopIB induces accumulation of positively supercoiled DNA in G1- and S-phase yeast cells. **a**, To eliminate the potential contribution of positive supercoils induced by replication forks, asynchronously growing haploid yeast cells are arrested in G1 phase with α -factor. **b**, Positive supercoils are generated ahead of advancing transcription bubbles, and compensatory negative supercoils are generated in their wake. Diverging transcription units bisect a plasmid into twin domains of positive and negative supercoils. If, in the presence of drug (left), positive supercoils are removed more slowly than negative supercoils, the plasmid should display an excess of positive supercoils. In the absence of drug (right), no asymmetry between the removal rates of positive supercoils and negative supercoils is observed (Supplementary Information VII) and no net change in linking number (Lk) is expected. **c**, Two-dimensional agarose gel analysis of the Lk distribution of 2 μ m plasmid topoisomers isolated from G1-phase *top1 Δ* , wild-type *TOP2* yeast cells, expressing plasmid-encoded human TopIB (TopIB) or vector control, and treated with camptothecin (CPT) or no drug for 15 or 60 min. Topoisomer resolution in two-dimensional agarose gels is discussed in Supplementary Information VI. The shift in topoisomer distribution is quantified by measuring the signal intensity of the portion of the arc to the right of the diagonal linear

protein, with vaccinia TopIB (which is insensitive to topotecan) or with a nick induced by a nicking enzyme (Supplementary Information V). As expected, these experiments show that catalytically active human TopIB is required to yield slow uncoiling, consistent with the *in vivo* activity and co-crystal structures of camptothecin analogues^{7,19,23}. We thus conclude that slow uncoiling represents a clear signature for a topotecan molecule bound to the TopIB–DNA complex.

Highly relevant for any consideration of the ‘collision’ model is the lifetime of the covalent complex. Consequently, we asked over what timescale topotecan remains bound, and thus over what timescale TopIB remains covalently linked to the DNA. Figure 2a shows a time trace indicating the times at which topotecan enters (t_{start}) and exits (t_{end}) the covalent complex. The assignment of t_{start} occurs after a plateau is interrupted by a fast rise in DNA extension as a result of DNA cleavage by TopIB, followed by the appearance of slow uncoiling at t_{start} , which we attribute to a topotecan molecule entering a TopIB-linked nick. Figure 2b–e shows a collection of traces illustrating the assignment of t_{start} and t_{end} . The majority of events observed (~70%) resemble those in Fig. 2b, c. In Fig. 2b, at ~340 s, the magnets are rotated for a substantial number of turns on two occasions (the interruption of the DNA extension signal at 334 and 338 s signifies rotation), but no plectonemes are introduced. This situation corresponds to a nicked DNA resulting from the formation of the covalent complex. At t_{start} a topotecan molecule enters the covalent complex; at this stage, slow uncoiling is observed. Figure 2c shows the abrupt transition from slow to fast supercoil removal, prompting the designation of t_{end} . The bound time Δt is defined as $t_{\text{end}} - t_{\text{start}}$ and reflects the time supercoils are enzymatically removed at a constant slow rate (shaded box, Fig. 2a). Multiple mechanical recoiling events are often required to measure Δt (four shown in Fig. 2a). Figure 2f shows the distribution of Δt , which has a mean of 121 ± 11 s ($n = 146$) that is unchanged both within the practical force range of the technique (Fig. 2f, inset) and by a tenfold change in topotecan concentration (Supplementary Information III). The latter indicates that the lifetime is unlikely to be overestimated by topotecan unbinding/rebinding during mechanical recoiling or by the presence of occasional very short-lived plateaus (Supplementary Information III). We conclude that, in the presence of topotecan, TopIB remains trapped on the DNA for at least 121 s. This timescale is about 400 times longer than the religation time in the absence of topotecan (Supplementary Information VI), the distribution of which has a most probable value of ~0.3 s (Fig. 2g), with a tail that extends to longer times (Fig. 2g inset). This provides quantitative support for a critical aspect of the collision model, namely the topotecan-induced increase in the lifetime of the covalent complex.

We next focused on the enzymatic uncoiling rate of positive versus negative supercoils. In the absence of topotecan, no significant difference was detected (Supplementary Information VII). However, we observe a clear and unexpected difference in supercoil removal rate in the presence of topotecan. Figure 2h shows that the uncoiling of positive supercoils (pink regions) is significantly slower than the uncoiling of negative supercoils (blue region). This asymmetry in uncoiling rates is quantified in Fig. 2i as the differential in rates, denoted $\Delta\omega$ and defined as $|\omega_+| - |\omega_-|$, where $|\omega_+|$ is the uncoiling rate of positive supercoils and $|\omega_-|$ is the uncoiling rate of negative supercoils. We find a mean $\Delta\omega$ of -8 ± 5 Hz ($n = 17$). The relative difference, $\Delta\omega/|\omega_+|$, is -2.4 ± 2.2 . The microscopic interactions responsible for the asymmetric rate of DNA uncoiling in the presence of topotecan were unforeseen by crystallographic studies, which only provide a static picture of crystalline TopIB.

DNA (purple arc) relative to the amount of label detected across the entire arc (purple numbers). **d**, Topoisomer distribution of 2 μ m plasmid in *top1 Δ* , wild-type *TOP2* yeast cells released into S phase. Here, too, accumulation of positive supercoils is observed only when cells express TopIB and are treated with camptothecin.

To investigate the biological ramifications of the asymmetric hindrance of DNA uncoiling induced by topotecan, we asked if this bias was evident in drug-treated yeast cells. Our single-molecule observations led us to posit that positive supercoils would accumulate during cellular processes that induce DNA supercoiling, such as transcription^{24,25}. Thus, we treat G1-phase-arrested *top1Δ TOP2* yeast cells (Fig. 3a) expressing low levels of plasmid-encoded human TopIB with camptothecin (Fig. 3b, left) or no drug (Fig. 3b, right). Arresting cells in G1 phase restricted the analysis to transcription-induced DNA supercoiling, and camptothecin was used to avoid drug transporter efflux of topotecan¹¹. Two-dimensional gel electrophoresis (Supplementary Information VIII) resolves the distribution of purified 2 μ m plasmid DNA topoisomers (Fig. 3c). In the absence of drug (Fig. 3c, upper row), an expected bias in the plasmid topoisomer distribution towards negative supercoils is observed, caused by DNA wrapping around the histone core in nucleosomes. However, camptothecin treatment of TopIB-expressing cells in G1 phase induces a remarkable skewing of the plasmid topoisomer distribution towards positive supercoiling (Fig. 3c, purple numbers quantify the skewing). No alteration in linking number was induced by camptothecin treatment of vector control cells that lack TopIB (Fig. 3c) or of cells expressing the catalytically inactive human Top1(Y723F) mutant (Supplementary Information IX); in both experiments the topoisomer distribution appears as in the case of no treatment. Combined with our single-molecule observations, these *in vivo* findings support a model in which the positive supercoils that accumulate ahead of the transcription bubble are removed less effectively by TopIB in the presence of drug than negative supercoils, which leads to a persistent overwinding of DNA in the absence of DNA replication in G1 phase.

Camptothecin toxicity is linked to on-going DNA replication in S phase^{8,9}, where positive supercoils accumulate ahead of the replication fork and are removed by TopIB. To study the effect of camptothecin treatment on TopIB activity during S phase, cells arrested in G1 phase were allowed to synchronously enter S phase by removal of α -factor from the culture medium, after which the experiment was performed as above (for cell cycle distribution and cell viability data, see Supplementary Information X). An analysis of 2 μ m DNA topoisomers purified from S-phase cells (Fig. 3d) yields similar results to that observed in the G1-phase experiments: positive-supercoil accumulation occurs only in TopIB-expressing cells and only in the presence of camptothecin.

In G1- and S-phase, the accumulation of positive supercoils was slightly enhanced in yeast cells expressing a thermosensitive Top2

(Supplementary Information XI). Although Top2 is essential to maintain cell viability²⁶, these observations indicate that the positive supercoils induced by camptothecin poisoning of TopIB are not effectively resolved by Top2. Thus, camptothecin-induced accumulation of positive supercoils occurs in a background of other topoisomerases, is not confined to a single phase of the cell cycle, and may derive from the tracking of a variety of complexes along the DNA duplex, including the transcription and replication machinery.

Our single-molecule data provide a direct determination of the vastly increased lifetime of the covalent TopIB–DNA complex stabilized by topotecan. However, we also observed a second consequence of drug binding: topotecan hinders DNA uncoiling by TopIB such that negative supercoils are relaxed more readily than positive supercoils. These findings provide a compelling argument for an intriguing, previously unconsidered, mechanism of drug-induced cell death (Fig. 4a, left). We propose that positive supercoils generated ahead of the fork (which cannot be efficiently removed by drug-bound TopIB or wild-type levels of Top2) may hamper fork progression. This stalling of the replication machinery could result in fork collapse and the formation of potentially lethal DNA lesions that induce cell death.

Implicit in this model is that the ability of camptothecins to selectively hinder the uncoiling of positively supercoiled DNA by TopIB coincides with drug cytotoxicity. Therefore, if the accumulation of positive supercoils is predictive of cellular response to this class of chemotherapeutics, then cells expressing a catalytically active, yet camptothecin-resistant, TopIB mutant enzyme would fail to accumulate positively supercoiled plasmid DNA in the presence of camptothecin. To test this critical aspect of the model, *top1Δ* yeast cells were transformed with a plasmid expressing either wild-type human TopIB or the catalytically active, camptothecin-resistant mutant Top1(G365C), which harbours a single amino acid substitution of Cys for Gly 365 (ref. 27). Figure 4b shows the 2 μ m DNA topoisomer distribution from asynchronously growing cells containing wild-type TopIB (left) or Top1(G365C) (right). This mutant failed to induce positive supercoil accumulation in the presence of camptothecin, in full agreement with our proposed mechanism that positive supercoils are involved in TopIB-mediated and camptothecin-induced cell death. Thus, a single mutation in an otherwise identical cellular background provided a stringent test of the specificity of TopIB–camptothecin interactions in causing the accumulation of positive supercoils associated with camptothecin-induced cell death.

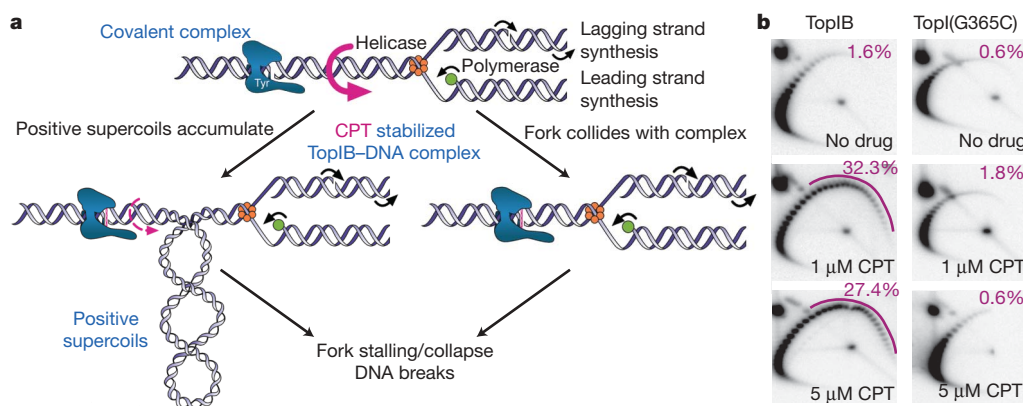


Figure 4 | Mechanism of camptothecin-induced toxicity derives from TopIB-dependent accumulation of positive supercoils. **a**, Representation of distinct mechanisms potentially underlying camptothecin-induced cell death. **A** (diagrammatically simplified) replication fork generates positive supercoils in the DNA, which are removed (magenta arrow) by TopIB (blue). In the presence of camptothecin, fork stalling and collapse have been predicted to result from the physical collision of the advancing replication complex with the drug-stabilized TopIB–DNA covalent complex (right).

Our data suggest a second scenario, in which fork progression and integrity are indirectly impaired by unresolved positive supercoils (left). **b**, Analysis of 2 μ m DNA topoisomer distribution in asynchronous cultures of *top1Δ* cells expressing either human wild-type TopIB or a catalytically active, camptothecin-resistant Top1(G365C) mutant. These were treated with indicated concentrations of camptothecin for 15 min. The accumulation of positive supercoils is only observed in drug-treated cells expressing wild-type TopIB.

METHODS

DNA constructs. Single-molecule experiments²⁸ used λ -phage DNA (48 kilobases (kb) or 16 μ m contour length) or half-length λ -phage DNA (24 kb or 8 μ m contour length). Measurements of the religation time were performed on shorter DNA (8 kb or 2.7 μ m contour length) to decrease the magnitude of the brownian fluctuations of the bead and to increase the time resolution.

Enzyme and buffers. Full-length human TopIB and catalytically inactive TOP1(Y723F), each containing an amino-terminal Flag epitope, were partially purified from galactose-induced cultures of EKY3 *top1 Δ* yeast cells²³. To obtain homogenous protein preparations, TopIB fractions, bound to an anti-Flag M2 affinity gel and eluted with an excess of Flag peptide in Tris-buffered saline (50 mM Tris (pH 7.4), 150 mM KCl), were applied to a phosphocellulose column. Homogeneous TopIB was eluted in TEEG buffer (50 mM Tris (pH 7.4), 1 mM EDTA, 1 mM EGTA, 10% glycerol) plus 1.0 M KCl and protease inhibitors, diluted with 50% glycerol and stored at -20°C . TopIB activity was assayed in plasmid DNA relaxation reactions and protein integrity was assessed in immunoblots²³.

Strains and plasmids. *Saccharomyces cerevisiae* strains EKY2 (MAT α , *top1 Δ*) and EKY3 (MAT α , *top1 Δ*) have been described²⁹. For constitutive expression of TopIB or TOP1(G365C), the epitope-tagged human TOP1 sequences (*ehTOP1*) were excised from YCpGAL1-*ehTOP1* (ref. 23) or YCpGAL1-*top1*(G365C) (ref. 27), and cloned under the yeast TOP1 promoter in YCpScehTOP1•U or YCpScehTOP1(G365C)•U. The empty vector YCpSc•U served as a control.

In vivo assays and two-dimensional gel electrophoresis. Exponential cultures of MAT α cells, transformed with YCpScehTOP1•U or vector control, were arrested in G1 phase with α -factor, then either incubated with additional α -factor and 30 μ M camptothecin or 0.25% DMSO (v/v) (no drug), or washed by filtration and released into S phase with or without camptothecin. Asynchronous cultures, incubated with 1 or 5 μ M camptothecin for 15 min at 30°C , were treated as above. To assess the 2 μ m plasmid DNA topoisomer distribution, DNA isolated from cultures fixed with toluene/ethanol was resolved by two-dimensional gel electrophoresis and subjected to Southern blotting^{29,30}.

Received 6 January; accepted 15 May 2007.

Published online 24 June 2007.

1. Champoux, J. J. DNA topoisomerases: structure, function, and mechanism. *Annu. Rev. Biochem.* **70**, 369–413 (2001).
2. Li, T. K. & Liu, L. F. Tumor cell death induced by topoisomerase-targeting drugs. *Annu. Rev. Pharmacol. Toxicol.* **41**, 53–77 (2001).
3. Pommier, Y. Topoisomerase I inhibitors: camptothecins and beyond. *Nature Rev. Cancer* **6**, 789–802 (2006).
4. Wang, J. C. Cellular roles of DNA topoisomerases: a molecular perspective. *Nature Rev. Mol. Cell Biol.* **3**, 430–440 (2002).
5. Hsiang, Y. H., Hertzberg, R., Hecht, S. & Liu, L. F. Camptothecin induces protein-linked DNA breaks via mammalian DNA topoisomerase I. *J. Biol. Chem.* **260**, 14873–14878 (1985).
6. Porter, S. E. & Champoux, J. J. The basis for camptothecin enhancement of DNA breakage by eukaryotic topoisomerase I. *Nucleic Acids Res.* **17**, 8521–8532 (1989).
7. Staker, B. L. *et al.* The mechanism of topoisomerase I poisoning by a camptothecin analog. *Proc. Natl Acad. Sci. USA* **99**, 15387–15392 (2002).
8. Holm, C., Covey, J. M., Kerrigan, D. & Pommier, Y. Differential requirement of DNA replication for the cytotoxicity of DNA topoisomerase I and II inhibitors in Chinese hamster DC3F cells. *Cancer Res.* **49**, 6365–6368 (1989).
9. Hsiang, Y. H., Lihou, M. G. & Liu, L. F. Arrest of replication forks by drug-stabilized topoisomerase I–DNA cleavable complexes as a mechanism of cell killing by camptothecin. *Cancer Res.* **49**, 5077–5082 (1989).
10. Corbett, K. D. & Berger, J. M. Structure, molecular mechanisms, and evolutionary relationships in DNA topoisomerases. *Annu. Rev. Biophys. Biomol. Struct.* **33**, 95–118 (2004).
11. Bjornsti, M. A. Cancer therapeutics in yeast. *Cancer Cell* **2**, 267–273 (2002).
12. Minsky, B. D. Combined-modality therapy of rectal cancer with irinotecan-based regimens. *Oncology (Huntingt.)* **18**, 49–55 (2004).
13. Rodriguez-Galindo, C. *et al.* Clinical use of topoisomerase I inhibitors in anticancer treatment. *Med. Pediatr. Oncol.* **35**, 385–402 (2000).
14. Stewart, D. J. Topotecan in the first-line treatment of small cell lung cancer. *Oncologist* **9** (Suppl 6), 33–42 (2004).
15. Sekiguchi, J. & Shuman, S. Vaccinia topoisomerase binds circumferentially to DNA. *J. Biol. Chem.* **269**, 31731–31734 (1994).
16. Redinbo, M. R., Stewart, L., Kuhn, P., Champoux, J. J. & Hol, W. G. Crystal structures of human topoisomerase I in covalent and noncovalent complexes with DNA. *Science* **279**, 1504–1513 (1998).
17. Koster, D. A., Croquette, V., Dekker, C., Shuman, S. & Dekker, N. H. Friction and torque govern the relaxation of DNA supercoils by eukaryotic topoisomerase IB. *Nature* **434**, 671–674 (2005).
18. Stivers, J. T., Harris, T. K. & Mildvan, A. S. Vaccinia DNA topoisomerase I: evidence supporting a free rotation mechanism for DNA supercoil relaxation. *Biochemistry* **36**, 5212–5222 (1997).
19. Chrencik, J. E. *et al.* Mechanisms of camptothecin resistance by human topoisomerase I mutations. *J. Mol. Biol.* **339**, 773–784 (2004).
20. Strick, T., Allemand, J., Croquette, V. & Bensimon, D. Twisting and stretching single DNA molecules. *Prog. Biophys. Mol. Biol.* **74**, 115–140 (2000).
21. Champoux, J. J. & Aronoff, R. The effects of camptothecin on the reaction and the specificity of the wheat germ type I topoisomerase. *J. Biol. Chem.* **264**, 1010–1015 (1989).
22. Stewart, L., Ireton, G. C. & Champoux, J. J. A functional linker in human topoisomerase I is required for maximum sensitivity to camptothecin in a DNA relaxation assay. *J. Biol. Chem.* **274**, 32950–32960 (1999).
23. Woo, M. H., Vance, J. R., Marcos, A. R., Bailly, C. & Bjornsti, M. A. Active site mutations in DNA topoisomerase I distinguish the cytotoxic activities of camptothecin and the indolocarbazole, rebeccamycin. *J. Biol. Chem.* **277**, 3813–3822 (2002).
24. Giaever, G. N. & Wang, J. C. Supercoiling of intracellular DNA can occur in eukaryotic cells. *Cell* **55**, 849–856 (1988).
25. Liu, L. F. & Wang, J. C. Supercoiling of the DNA template during transcription. *Proc. Natl Acad. Sci. USA* **84**, 7024–7027 (1987).
26. Goto, T. & Wang, J. C. Yeast DNA topoisomerase II is encoded by a single-copy, essential gene. *Cell* **36**, 1073–1080 (1984).
27. Woo, M. H. *et al.* Locking the DNA topoisomerase I protein clamp inhibits DNA rotation and induces cell lethality. *Proc. Natl Acad. Sci. USA* **100**, 13767–13772 (2003).
28. Strick, T. R., Allemand, J. F., Bensimon, D., Bensimon, A. & Croquette, V. The elasticity of a single supercoiled DNA molecule. *Science* **271**, 1835–1837 (1996).
29. Hann, C. *et al.* Increased camptothecin toxicity induced in mammalian cells expressing *Saccharomyces cerevisiae* DNA topoisomerase I. *J. Biol. Chem.* **273**, 8425–8433 (1998).
30. Megonigal, M. D., Fertala, J. & Bjornsti, M. A. Alterations in the catalytic activity of yeast DNA topoisomerase I result in cell cycle arrest and cell death. *J. Biol. Chem.* **272**, 12801–12808 (1997).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank K. Besteman and U. Keyser for discussions, S. Hage and Y.-H. Chien for DNA preparation, R. van Waardenburg for help with PyMol, and C. Dekker for critical reading and general support. We thank FOM, NWO, NIH and ALSAC for financial support.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to N.H.D. (n.h.dekker@tudelft.nl).

naturejobs

**JOBS OF
THE WEEK**

Life in a government laboratory can be pretty good. For a start, the money tends to be above average, and job security is usually fairly assured. On top of that, employees are free from the everyday administrative and teaching duties that they are saddled with at universities. All of this makes a career at a government lab sound quite appealing — but the institutions do vary from country to country, and even those with a high profile can sometimes find themselves on the back foot.

Over the next two weeks, *Naturejobs* takes a look at two government labs — the Los Alamos National Laboratory in New Mexico (see page 220) and the institutes run by the UK Medical Research Council (MRC; see next week's issue). Although there are many similarities between the labs, including a strong history of landmark discoveries, there are also significant differences. The size of research groups tends to be smaller in Britain, for example, and national security concerns have had serious repercussions at Los Alamos.

Indeed, as well as the need to adapt to the changing face of nuclear-weapons research, Los Alamos has found itself embarrassed by a series of security lapses. These, in turn, have led to changes in management, revamped procedures and, ultimately, a degree of staff attrition. Researchers at the lab may not have teaching duties, but they do have a veritable mountain of forms to fill out. But despite its recent problems, Los Alamos is still lauded for its interdisciplinary, cutting-edge research, high number of potential collaborators and ample resources — so it should be possible for the management and staff to restore its lost lustre.

But perhaps the major challenge facing both Los Alamos and the MRC is the quest for talent. Although government labs offer good salaries, the private sector — whether pharmaceutical companies, biotechnology firms or Internet companies — often offers even better. Prestigious labs such as Los Alamos or those run by the MRC have their perks and attractions, but even they must strive constantly to find the people and research that will keep their decades-old reputations intact.

Gene Russo, acting editor of *Naturejobs*

CONTACTS

Acting Editor: Gene Russo

European Head Office, London

The Macmillan Building,
4 Crinan Street,
London N1 9XW, UK
Tel: +44 (0) 20 7843 4961
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

European Sales Manager:

Andy Douglas (4975)
e-mail: a.douglas@nature.com
**Business Development
Manager:**
Amelie Pequignot (4974)
e-mail: a.pequignot@nature.com

Natureevents:

Claudia Paulsen Young
(+44 (0) 20 7014 4015)
e-mail: c.paulsenyoung@nature.com

France/Switzerland/Belgium:

Muriel Lestringuez (4994)

Southwest UK/RoW:

Nils Moeller (4953)

Scandinavia/Spain/Portugal/Italy:

Evelina Rubio-Hakansson (4973)

Northeast UK/Ireland:

Matthew Ward (+44 (0) 20 7014 4059)

North Germany/The Netherlands:

Reya Silao (4970)

South Germany/Austria:

Hildi Rowland (+44 (0) 20 7014 4084)

Advertising Production Manager:

Stephen Russell
To send materials use London
address above.

Tel: +44 (0) 20 7843 4816

Fax: +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Jasmine Myer

US Head Office, New York

75 Varick Street, 9th Floor,
New York, NY 10013-1917
Tel: +1 800 989 7718
Fax: +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Manager:

Peter Bless

Japan Head Office, Tokyo

Chiyoda Building,
2-37 Ichigayatamachi,
Shinjuku-ku, Tokyo 162-0843
Tel: +81 3 3267 8751
Fax: +81 3 3267 8746

Asia-Pacific Sales Manager:

Ayako Watanabe
Tel: +81-3-3267-8765
e-mail: a.watanabe@natureasia.com



J. RAEDLE/GETTY

FIGHTING BACK

Surrounded by colourful mesas and wide swaths of desert grassland in New Mexico, Los Alamos National Laboratory (LANL) boasts stunning views. Indeed, that beauty is one reason why Robert Oppenheimer pushed to house one of the nation's first nuclear-weapons labs there in 1943; the location's isolation was another. But things have changed since then. The gorgeous surroundings are still cited as a reason to work there, but a changing weapons mission, security scandals and frustrations with the lab's administration have left some wondering if LANL really is an ideal place to work, after all.

In its favour, the presence of world-class experts, a chance to do science potentially relevant to the nation's security and a lab-wide commitment to maintain current levels of basic research continue to bring researchers to Los Alamos. Working at a US national lab — of which there are some 30 supported by the Department of Energy — is an attractive prospect. These labs, although often run by an outside university or business, nevertheless benefit from a steady stream of government money. Scientists still have to seek grants, and they must, to varying degrees, perform research that fits in with government goals. But the sheer diversity of fields yields cross-disciplinary opportunities that a university can rarely offer, and the support of a large institution is often helpful for a researcher who is just starting out.

Still, Los Alamos has had some very public problems in the past five years, which has left its attractiveness in question. The lab is still going through a system-wide administration adjustment. Everyone admits it has been quite rough at times.

First, there were the security scandals in 2000, 2003 and 2004 — splashed all over the news — when classified information on disks and drives went missing. In July 2004, then-director Pete Nanos shut down lab operations in response to what he described as “wilful flouting of the rules” by employees. The unprecedented tactic resulted in the resignation of employees such as Daniel James, who had worked at Los Alamos since 1994, and now teaches at the

Emerging from scandal and reorganization, Los Alamos National Laboratory hopes to retain talent and remain at the cutting edge.

Karen Fox reports.



“Solving problems that change society is what attracted great scientists, and that is the thing we still offer.”
— Terry Wallace



University of Toronto. “Nanos started

screaming at us and shut the lab down, but we had to come to work and charge our salary to our respective grants,” says James. “We had to sit at our desks and read manuals on safe work practices. I’m a theoretical physicist — ‘safe work practices’ means avoiding paper cuts and carpal tunnel syndrome.”

Several years later, Terry Wallace, LANL's associate director of science, technology and engineering, says that 50 researchers (out of more than 9,000) left because of Nanos. According to Wallace, many other employees agreed that Nanos's leadership was problematic and also complained — but sought Nanos's departure rather than leaving themselves.

Corporate concerns

Nanos left in May 2005. But the US government still had security concerns, leading to a change in control from the University of California to Los Alamos National Security (LANS). The University of California remains part of the LANS team, but the days of a hands-off administration are gone, with an increased emphasis on security, safety and profit. Some LANL researchers, especially the long-time employees, find the more business-like, corporate environment distasteful. Many at the lab feared that when LANS came on board, people would retire in droves, but that hasn't happened.

“We haven't seen any change in our rate of retirement,” says Wallace. “Scientists don't really like to retire. They continue to stay on, and that's true here too.” But Wallace acknowledges that the lab is still going through a painful settling-in process with the new management. The current system gives individual labs financial rewards based on how safety-compliant

the researchers are, and how many security issues they have. "Often that doesn't have to do with science," he says. "To be the best science lab, these things can be in conflict, and reaching equilibrium is going to take a few more years."

And, of course, there's the budget. After decades of growth — growing from 3,600 employees in 1975 to well over double that today — Los Alamos, along with many other Department of Energy labs, is cutting back on recruitment in response to money restrictions. In the case of Los Alamos, the lab has the additional expense of paying the new LANS team. LANS charges a higher management fee than did the University of California, and, because LANS is a corporation, LANL has higher taxes. With 9,050 employees (200 fewer than last year), the lab is looking to cut back at least another 300 people. There are fewer new hires — almost none who weren't postdocs at LANL — and the administration is more involved in recruitment, according to Wallace.

Most of the researchers who focus on weapons are close to retirement age, leaving LANL with the dilemma of how to retain an important body of knowledge while recruiting new talent. It is not offering retirement incentives; however, there will undoubtedly be more retiring in the next few years, making room for new postdocs. Indeed, postdoctoral hiring is the one area that has not decreased, despite the shrinking budget — that number will stay at 350 at any given time, the largest postdoc programme of the US physics-based labs. And a postdoctoral fellowship continues to be the best route to a long-term position at LANL. About one-third of the postdocs are currently chosen for full-time positions, although that number is predicted to fall a little.

David Chavez, who converted from a postdoc to a full-time organic chemistry researcher in April 2006, acknowledges a steep learning curve when he made the switch. The various administrative responsibilities (such as filing safety and security reports) are demanding, but he values the chance to study national security problems with a societal impact. "Because of the broad range of science that goes on here, my



"Because of the broad range of science that goes on here, my intellectual capacity is always being pushed."
— David Chavez

intellectual capacity is always being pushed," he says.

The postdocs certainly seem to want to stay. Laura Stonehill, who began her postdoc in dark-matter research at LANL in 2005, is currently seeking security clearance, hoping it will increase her chances of a permanent job. The lab's diverse research areas enable her and her husband, also an employee, to find interesting work, and Stonehill loves the intellectual stimulation of such a large laboratory. "The transition to LANS has definitely been hard on everybody," she says, "but most of the postdocs are extremely happy." They like the ample resources and the large number of experts, not to mention the gorgeous outdoors and "cool-looking mesas", she says.

National security — or 'behind the fence' — projects require an adjustment. Postdoc work has more of a basic research focus: Chavez is now careful to make sure that his work on energetic materials, used in propellants and explosives, is relevant to national security. Those not interested in such goal-oriented research may be better off at a university.

"I got the impression Los Alamos was really down on things that weren't directly associated with the primary mission of nuclear weapons," James says. "They say differently, but actions speak louder than words." LANL has decided not to reduce the roughly 6% of lab funds set aside for basic research. But national security is still the major focus. It is a weapons lab, after all.

New focus

Nevertheless, the programmatic research is less weapons-based than it once was. The Nuclear-Test-Ban Treaty (initiated in 1996, and now with 177 state signatories) put a stop to nuclear-weapons-testing explosions, although computer-modelling research continues. There is also a new focus on nonproliferation research. This can mean anything from the more theoretical, such as studying cosmic background radiation in an attempt to track uranium or plutonium movements anywhere in the world by noting radiation footprints, to the more applied, such as creating detectors for nuclear materials.

Other emerging areas include biophysics and physics beyond the current model of particles, called the standard model. Wallace also emphasizes the need for information-science technology skills, regardless of speciality. Taking complicated data sets and turning them into useful information is crucial for almost every aspect of LANL's work. Indeed, it faces stiff recruiting competition from companies, such as Google, that pay more. But LANL, says Wallace, offers an environment where communal teams of scientists solve problems that change society. It's what lures young researchers to Los Alamos, he says. "That is what attracted great scientists such as [Hans] Bethe and [Richard] Feynman, and that is the thing we still offer," he says.

For many, the joy of working on applied research within such a big intellectual enclave is greater than the frustrations of living through a period of administrative change at the lab. "There's a joke that the postdocs are the only ones who can actually do anything, because the rest of us are filling out forms," says staff member Michelle Espy, who has worked at LANL off and on since starting graduate school. "But I remain optimistic. The lab is going to remain a world-class place to work. I just bought a new house — I'm planning to stay."

Karen Fox is a science writer based in Washington DC.



Fenced in: administration is less hands-off following a series of security scandals at LANL.

MOVERS

France Córdoba, president, Purdue University, West Lafayette, Indiana



2002-07: Chancellor, University of California, Riverside

1996-2002: Vice-chancellor for research and professor of physics, University of California, Santa Barbara

1993-96: Chief scientist, NASA, Washington DC

The career of France Córdoba, the new president of Purdue University in West Lafayette, Indiana, was hardly the product of planning. Lacking scientific mentors, she majored in English at Stanford. She pursued anthropology, journalism and education, but wasn't satisfied with any of them.

A television special on neutron stars, after Neil Armstrong's historic visit to the Moon, changed her life by sparking an interest in astrophysics. She contacted the scientists at the Massachusetts Institute of Technology who were featured in the programme and got a summer job in their lab — which led to a graduate studentship in astrophysics at the California Institute of Technology, and on to a PhD.

Córdoba was attracted to high-energy astrophysics because the field was ripe for discovery. "I was in the vanguard of a small group of students proposing to do multi-wavelength observations to observe cosmic bodies of interest," she says. She then made the seminal discovery — soft X-ray pulsations from a class of close binary stars — that garnered her multiple job offers.

A permanent job at Los Alamos National Laboratory gave her ten years to start her career without having to move on to the next postdoc. Eager to be around students once more, she then accepted an offer to head Pennsylvania State University's department of astronomy and astrophysics. NASA administrator Dan Goldin would eventually select her to be the agency's first female chief scientist.

Goldin not only wanted to diversify the predominantly male establishment; he also wanted Córdoba to refocus the scientific agenda on fundamental questions, including the origin and evolution of the cosmos and a search for Earth-like planets. "As a result, we had a scientific explosion at NASA — particularly in astrophysics," says Goldin. He also asked Córdoba to make links among physics, chemistry and biology researchers at NASA, develop relationships with other agencies and steer NASA's international leadership in astrophysics. She delivered on every level, he says.

She has since overseen two University of California research agendas: as vice-chancellor at Santa Barbara and chancellor at Riverside. At Riverside she did groundwork for California's first new medical school in 45 years.

"I like complexity and thrive on stress — it keeps me energized," says Córdoba. She is now planning to extend Purdue's mission to include energy security, agricultural sustainability and information technology. "She's going to take it to the next level," says Goldin. ■

Virginia Gewin

NETWORKS & SUPPORT

Engineering a place for women

In Britain, only 15–20% of women with a degree in science, engineering and technology (SET) fields are working in a SET occupation. Some 50,000 have left the sector. And everyday workplace culture may be one major barrier to their recruitment, retention and advancement.

Representatives of universities, professional institutions and industry gathered at a seminar at the Institute of Physics (IOP) in London in late June to learn what 'workplace culture' means, how to assess it and how to change it. The seminar was organized by the UK Resource Centre for Women in Science, Engineering and Technology (UKRC), which was set up in 2004 with government funding to assist employers and individual women to tackle the problems.

Wendy Faulkner, of the Science Studies Unit at the University of Edinburgh, told delegates that the ways in which different individuals come to belong to the group may affect whether women are accepted. She suggested that workplace culture included styles of interacting, topics of conversation, humour and social circles. Crucially, said Faulkner, facets of masculine culture (such as football discussions and macho attitudes) mesh well with the culture of SET workplaces. Anecdotes from women

she has interviewed suggested that they found it hard to fit in, whether because they couldn't take part in a joke or because they weren't seen as 'real' engineers. In some cases, male colleagues made inappropriate remarks. One woman said she wanted to report a culprit, but felt she had neither the confidence nor the support. Delegates at the meeting mutely mouthed recognition.

Forty companies have made changes after using the UKRC's culture analysis tool of detailed questionnaires for managers and staff. Both the IOP and the Royal Society of Chemistry have used 'site visits' to assess cultures in university departments. This resulted in the IOP's formulation of a code of practice for addressing gender issues, including "appointment, promotion and selection processes and procedures that encourage men and women to apply for academic posts at all levels". This code could be adapted by organizations in other disciplines, says Peter Main, director of education and training at the IOP.

Studying and assessing workplace cultures is a difficult task. But if it helps to retain more women in SET fields, it's worthwhile nonetheless. ■

Roger Livesey is a public-relations officer at UKRC.

POSTDOC JOURNAL

(Almost) in press

I am currently writing the first manuscript from my postdoctoral research. We're very close to submitting to a journal, and I must say that I'm looking forward to getting it out the door. The latest edits haven't been about the scientific content of the manuscript so much as reshaping the story we are telling, arranging the data and sculpting the text to make it more compelling.

Preparing a manuscript for publication requires more stamina than I'd expected. My doctoral adviser and I are still trying to publish the last part of my dissertation research: research that was completed months ago. Despite its frustrations, the peer-review process seems to have worked well, as two rejections have made that manuscript much stronger. Those rejections were based not on the science, but on our inability to tell the story clearly. We're hoping that simplified explanations and a slight change of focus will do the trick.

Good science, I think, is as much about lucid communication as about proper pipette technique. In an ideal world, there would be other, simpler ways to disseminate one's findings than laboriously writing journal articles. Maybe one day, self-assembling data will be downloaded directly into interested scientists' brains, bypassing the need to construct an elaborate tale on paper. Until that fiction becomes reality, I'd better keep working on my storytelling skills. ■

Peter Jordan is a visiting fellow at the National Institute of Diabetes and Digestive and Kidney Diseases in Bethesda, Maryland.

A piratical sabbatical

It's just too good to be true.

Ian Whates

You're good, you know that? Well yes, I do realize that's what you're paid to be.

No, really, I wasn't being a smart-ass... sorry. I just never realized it was going to be this realistic.

How it all started? At college; the usual vacation-time question: 'What do I do now?' Everybody's doing something; even my kid sister's gallivanting around the galaxy with her boyfriend. I wasn't going to sit around kicking my heels now, was I? Then I saw this ad on campus...

What? My parents? They're off on some pampered luxury cruise: auto-masseuse, stimulants-on-tap — the works.

Anyway, so I see this ad: Pirate Experience, the chance to be a buccaneer for a few weeks, how cool is that?

I took the virtual tour and I was hooked!

It was so well put together. We were all taken to this deserted corner of the old space port. Darkness, flickering lights — it already felt brimming over with skulduggery, even before we were in the air.

I didn't know any of the others, which was part of the adventure.

No, I didn't bother asking anyone. Pointless — I knew all my friends were fixed up already. I would have been too, if Marcy hadn't dumped me; the two-faced...

Okay, sorry. We boarded this rickety shuttle, and guess what they piped over the sound system as we took off? Old sea shanties: 'Yo-ho-ho and a bottle of rum'. Stewards even came round with mugs of grog. It was all so authentic.

We docked with a larger craft waiting in orbit — *The Queen Anne's Revenge*. What kind of crazy name for a ship is that?

Sorry, haven't a clue. I'm not very good with ships... a frigate maybe?

Anyway, we boarded the *Queen Anne* and were each allocated a bunk — not hammocks thank goodness, but still pretty crude.

Oh, I'd say 25, maybe 30 of us. I never bothered counting.

We were divided into groups of four or five and given a rota of duties. One morning it would be minor patch work and repairs — there were plenty of those needed — the next it might be swabbing the decks... I kid you not. If that ship ever had auto-clean, it had broken down long ago, or been switched off especially for our benefit.

The mornings were spent on chores, while the afternoons were reserved for more fun things: gunnery lessons in the simulator — I was particularly good at that — target practice with pistols in the ship's range, even sword-fighting and some hand-to-hand stuff. Never realized there was so much involved in being a



pirate, to be honest. It was exhausting but exhilarating.

In the evenings, after dinner, the lights would dim and we'd sit around a holofire sipping grog and bumboo from our black-jacks — that's a sort of cup — while members of the regular crew told tales of famous pirates and their exploits.

The crew were great; they had all the jargon and would roar at us to 'avast' and say 'ahoy' instead of hello, stuff like that. There was even the odd 'shiver me timbers'.

And then, to cap it all, they 'discovered' a ship — a big ocean liner, ripe for plundering. We knew what that meant: time for some real piracy. Sirens sounded and everyone ran around. I was assigned to gunnery, as I'd done so well on the simulators.

To be honest, the piracy bit wasn't as much fun as I'd hoped. It was mainly just waiting around. I wasn't involved in actually breaching the cruise-ship's hull — I hadn't done that well at the simulators, apparently — but was given the job

of mopping up. So when two shuttles launched from the target after it had been breached, I shot them down.

Three shots, that's all it took, and I hit them both. Big explosions. Wham!

Of course I'm proud. It was some hot shooting and it's not as if anyone actually died, after all.

That's ridiculous. What reason would I have for *really* plundering an ocean liner?

Money? Come on, have you any idea how rich my parents are?

Ha, ha, that's a good one — no, of course I'm not trying to bribe you.

Really, Officer, I'm not trying to bribe you.

After the raid? Well, the crew were in fine spirits — extra grog all round and lots of singing.

Next morning there were a few sore heads and dodgy stomachs, I can tell you. But instead of our chores we were given these two sealed chests...

Yes, the ones you found us with. We were put ashore at some backwater planet and told to wait, that someone would be along shortly to show us where to bury the treasure.

Well, we waited and waited, until you lot showed up and took us all into custody. I have to say, the way you came screaming down in that cruiser with all those flashing lights and gleaming insignia was quite something — very convincing. The most authentic part so far.

I realize you've done it many times before, that this is just another part of the Experience, but even so...

That's ridiculous. My folks are major wealthy. I've no motive for pulling something like this in the real world.

What do you mean 'inheritance'?

The ship we hit? I can't remember. If they told us the name, I missed it.

Golden Star? Sorry, means nothing to me.

My parents' cruise ship? Oh that's good, that's really good. You had me going there for a minute.

Lucky for me I know this is all part of the package. None of this is real, right?

Right?

During the past year, seven of Ian's stories have appeared in various books and magazines, including *Nature*, the anthologies *Glorifying Terrorism* and BSFA-winning *Time Pieces* — which Ian edited and published.

JACEY